

# ON THE ESTIMATION OF INTEGRATED COVARIANCE MATRICES OF HIGH DIMENSIONAL DIFFUSION PROCESSES<sup>1</sup>

BY XINGHUA ZHENG AND YINGYING LI

*Hong Kong University of Science and Technology*

We consider the estimation of integrated covariance (ICV) matrices of high dimensional diffusion processes based on high frequency observations. We start by studying the most commonly used estimator, the *realized covariance* (RCV) matrix. We show that in the high dimensional case when the dimension  $p$  and the observation frequency  $n$  grow in the same rate, the limiting spectral distribution (LSD) of RCV depends on the covolatility process *not only through the targeting ICV, but also on how the covolatility process varies in time*. We establish a Marčenko–Pastur type theorem for weighted sample covariance matrices, based on which we obtain a Marčenko–Pastur type theorem for RCV for a class  $\mathcal{C}$  of diffusion processes. The results explicitly demonstrate how the time variability of the covolatility process affects the LSD of RCV. We further propose an alternative estimator, the *time-variation adjusted realized covariance* (TVARCV) matrix. We show that for processes in class  $\mathcal{C}$ , the TVARCV possesses the desirable property that its LSD depends solely on that of the targeting ICV through the Marčenko–Pastur equation, and hence, in particular, the TVARCV can be used to recover the empirical spectral distribution of the ICV by using existing algorithms.

## 1. Introduction.

1.1. *Background.* Diffusion processes are widely used to model financial asset price processes. For example, suppose that we have multiple stocks, say,  $p$  stocks whose price processes are denoted by  $S_t^{(j)}$  for  $j = 1, \dots, p$ , and  $X_t^{(j)} := \log S_t^{(j)}$  are the log price processes. Let  $\mathbf{X}_t = (X_t^{(1)}, \dots, X_t^{(p)})^T$ . Then

---

Received May 2010; revised October 2011.

<sup>1</sup>Supported in part by DAG (HKUST) and GRF 606811 of the HKSAR.

*AMS 2000 subject classifications.* Primary 62H12; secondary 62G99; tertiary 60F15.

*Key words and phrases.* High dimension, high frequency, integrated covariance matrix, Marčenko–Pastur equation, weighted sample covariance matrix, realized covariance matrix.

This is an electronic reprint of the original article published by the Institute of Mathematical Statistics in *The Annals of Statistics*, 2011, Vol. 39, No. 6, 3121–3151. This reprint differs from the original in pagination and typographic detail.

a widely used model for  $\mathbf{X}_t$  is [see, e.g., Definition 1 in Barndorff-Nielsen and Shephard (2004)]

$$(1.1) \quad d\mathbf{X}_t = \boldsymbol{\mu}_t dt + \Theta_t d\mathbf{W}_t,$$

where,  $\boldsymbol{\mu}_t = (\mu_t^{(1)}, \dots, \mu_t^{(p)})^T$  is a  $p$ -dimensional drift process;  $\Theta_t$  is a  $p \times p$  matrix for any  $t$ , and is called the (instantaneous) *covolatility process*; and  $\mathbf{W}_t$  is a  $p$ -dimensional standard Brownian motion.

The *integrated covariance* (ICV) matrix

$$\Sigma_p := \int_0^1 \Theta_t \Theta_t^T dt$$

is of great interest in financial applications, which in the one dimensional case is known as the *integrated volatility*. A widely used estimator of the ICV matrix is the so-called *realized covariance* (RCV) matrix, which is defined as follows. Assume that we can observe the processes  $X_t^{(j)}$ 's at high frequency synchronously, say, at time points  $\tau_{n,\ell}$ :

$$X_{\tau_{n,\ell}}^{(j)} (= \log S_{\tau_{n,\ell}}^{(j)}), \quad \ell = 0, 1, \dots, n, j = 1, \dots, p,$$

then the RCV matrix is defined as

$$(1.2) \quad \Sigma_p^{\text{RCV}} := \sum_{\ell=1}^n \Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T$$

where  $\Delta \mathbf{X}_\ell = \begin{pmatrix} \Delta X_\ell^{(1)} \\ \vdots \\ \Delta X_\ell^{(p)} \end{pmatrix} := \begin{pmatrix} X_{\tau_{n,\ell}}^{(1)} - X_{\tau_{n,\ell-1}}^{(1)} \\ \vdots \\ X_{\tau_{n,\ell}}^{(p)} - X_{\tau_{n,\ell-1}}^{(p)} \end{pmatrix}.$

In the one dimensional case, the RCV matrix reduces to the *realized volatility*. Thanks to its nice convergence to the ICV matrix as the observation frequency  $n$  goes to infinity [see Jacod and Protter (1998)], the RCV matrix is highly appreciated in both academic research and practical applications.

**REMARK 1.** The tick-by-tick data are usually not observed synchronously, and moreover are contaminated by market microstructure noise. On sparsely sampled data (e.g., 5-minute data for some highly liquid assets, or subsample from data synchronized by refresh times [Barndorff-Nielsen et al. (2011)]), the theory in this paper should be readily applicable, just as one can use the realized volatility based on sparsely sampled data to estimate the integrated volatility; see, for example, Andersen et al. (2001).

**1.2. Large dimensional random matrix theory (LDRMT).** Having a good estimate of the ICV matrix  $\Sigma_p$ , in particular, its spectrum (i.e., its set of eigenvalues  $\{\lambda_j : j = 1, \dots, p\}$ ), is crucial in many applications such as prin-

principal component analysis and portfolio optimization (see, e.g., the pioneer work of Markowitz (1952, 1959) and a more recent work [Bai, Liu and Wong (2009)]). When the dimension  $p$  is high, it is more convenient to study, instead of the  $p$  eigenvalues  $\{\lambda_j : j = 1, \dots, p\}$ , the associated *empirical spectral distribution* (ESD)

$$F^{\Sigma_p}(x) := \frac{1}{p} \#\{j : \lambda_j \leq x\}, \quad x \in \mathbb{R}.$$

A naive estimator of the spectrum of the ICV matrix  $\Sigma_p$  is the spectrum of the RCV matrix  $\Sigma_p^{\text{RCV}}$ . In particular, one wishes that the ESD  $F^{\Sigma_p^{\text{RCV}}}$  of  $\Sigma_p^{\text{RCV}}$  would approximate  $F^{\Sigma_p}$  well when the frequency  $n$  is sufficiently high. From the large dimensional random matrix theory (LDRMT), we now understand quite well that in the high dimensional setting this good wish won't come true. For example, in the simplest case when the drift process is 0, covolatility process is constant, and observation times  $\tau_{n,\ell}$  are equally spaced, namely,  $\tau_{n,\ell} = \ell/n$ , we are in the setting of estimating the usual *covariance matrix* using the *sample covariance matrix*, given  $n$  i.i.d.  $p$ -dimensional observations  $(\Delta \mathbf{X}_\ell)_{\ell=1, \dots, n}$ . From LDRMT, we know that if  $p/n$  converges to a non-zero number and the ESD  $F^{\Sigma_p}$  of the true covariance matrix converges, then the ESD  $F^{\Sigma_p^{\text{RCV}}}$  of the sample covariance matrix also converges; see, for example, Marčenko and Pastur (1967), Yin (1986), Silverstein and Bai (1995) and Silverstein (1995). The relationship between the *limiting spectral distribution* (LSD) of  $\Sigma_p^{\text{RCV}}$  in this case and the LSD of  $\Sigma_p$  can be described by a Marčenko–Pastur equation through Stieltjes transforms, as follows.

PROPOSITION 1 [Theorem 1.1 of Silverstein (1995)]. *Assume on a common probability space:*

- (i) *for  $p = 1, 2, \dots$  and for  $1 \leq \ell \leq n$ ,  $\mathbf{Z}_\ell^{(p)} = (Z_\ell^{(p,j)})_{1 \leq j \leq p}$  with  $Z_\ell^{(p,j)}$  i.i.d. with mean 0 and variance 1;*
- (ii)  *$n = n(p)$  with  $y_n := p/n \rightarrow y > 0$  as  $p \rightarrow \infty$ ;*
- (iii)  *$\Sigma_p$  is a (possibly random) nonnegative definite  $p \times p$  matrix such that its ESD  $F^{\Sigma_p}$  converges almost surely in distribution to a probability distribution  $H$  on  $[0, \infty)$  as  $p \rightarrow \infty$ ;*
- (iv)  *$\Sigma_p$  and  $\mathbf{Z}_\ell^{(p)}$ 's are independent.*

Let  $\Sigma_p^{1/2}$  be the (nonnegative) square root matrix of  $\Sigma_p$  and  $S_p := 1/n \times \sum_{\ell=1}^n \Sigma_p^{1/2} \mathbf{Z}_\ell^{(p)} (\mathbf{Z}_\ell^{(p)})^T \Sigma_p^{1/2}$ . Then, almost surely, the ESD of  $S_p$  converges in distribution to a probability distribution  $F$ , which is determined by  $H$  in that its Stieltjes transform

$$m_F(z) := \int_{\lambda \in \mathbb{R}} \frac{1}{\lambda - z} dF(\lambda), \quad z \in \mathbb{C}_+ := \{z \in \mathbb{C} : \text{Im}(z) > 0\}$$

solves the equation

$$(1.3) \quad m_F(z) = \int_{\tau \in \mathbb{R}} \frac{1}{\tau(1 - y(1 + zm_F(z))) - z} dH(\tau).$$

In the special case when  $\Sigma_p = \sigma^2 I_{p \times p}$ , where  $I_{p \times p}$  is the  $p \times p$  identity matrix, the LSD  $F$  can be explicitly expressed as follows.

PROPOSITION 2 [see, e.g., Theorem 2.5 in Bai (1999)]. *Suppose that  $\mathbf{Z}_\ell^{(p)}$ 's are as in the previous proposition, and  $\Sigma_p = \sigma^2 I_{p \times p}$  for some  $\sigma^2 > 0$ . Then the LSD  $F$  has density*

$$p(x) = \frac{1}{2\pi\sigma^2 xy} \sqrt{(b-x)(x-a)} \quad \text{if } a \leq x \leq b,$$

and a point mass  $1 - 1/y$  at the origin if  $y > 1$ , where

$$(1.4) \quad a = a(y) = \sigma^2(1 - \sqrt{y})^2 \quad \text{and} \quad b = b(y) = \sigma^2(1 + \sqrt{y})^2.$$

The LSD  $F$  in this proposition is called the Marčenko–Pastur law with ratio index  $y$  and scale index  $\sigma^2$ , and will be denoted by  $\text{MP}^{(y, \sigma^2)}$  in this article.

**1.3. Back to the stochastic volatility case.** In practice, the covolatility process is typically not constant. For example, it is commonly observed that the stock intraday volatility tends to be U-shaped [see, e.g., Admati and Pfleiderer (1988), Andersen and Bollerslev (1997)] or exhibits some other patterns [see, e.g., Andersen and Bollerslev (1998)]. In this article, we shall allow them to be not only varying in time but also stochastic. Furthermore, we shall allow the observation times  $\tau_{n,\ell}$  to be random. These generalizations make our study to be different in nature from the LDRMT: in LDRMT the observations are i.i.d.; in our setting, the observations  $(\Delta \mathbf{X}_\ell)_{\ell=1, \dots, n}$  may, first, be dependant with each other, and second, have different distributions because (i) the covolatility process may vary over time, and (ii) the observation durations  $\Delta \tau_\ell := \tau_{n,\ell} - \tau_{n,\ell-1}$  may be different.

In general, for any time-varying covolatility process  $\Theta_t$ , we associate it with a constant covolatility process given by the square root of the ICV matrix

$$(1.5) \quad \Theta_t^0 := \sqrt{\int_0^1 \Theta_s \Theta_s^T ds} \quad \text{for all } t \in [0, 1].$$

Let  $\mathbf{X}_t^0$  be defined by replacing  $\Theta_t$  with the constant covolatility process  $\Theta_t^0$  (and replacing  $\mu_t$  with 0, and  $\mathbf{W}_t$  with another independent Brownian motion, if necessary) in (1.1). Observe that  $\mathbf{X}_t$  and  $\mathbf{X}_t^0$  share the same ICV

matrix at time 1. Based on  $\mathbf{X}_t^0$ , we have an associated RCV matrix

$$(1.6) \quad \Sigma_p^{\text{RCV}^0} = \sum_{\ell=1}^n \Delta \mathbf{X}_\ell^0 (\Delta \mathbf{X}_\ell^0)^T,$$

which is estimating the same ICV matrix as  $\Sigma_p^{\text{RCV}}$ .

Since  $\Sigma_p^{\text{RCV}}$  and  $\Sigma_p^{\text{RCV}^0}$  are based on the same estimation method and share the same targeting ICV matrix, it is desirable that their ESDs have similar properties. In particular, based on the results in LDRMT and the discussion about constant covolatility case in Section 1.2, we have the following property for  $\Sigma_p^{\text{RCV}^0}$ : if the ESD  $F^{\Sigma_p}$  converges, then so does  $F^{\Sigma_p^{\text{RCV}^0}}$ ; moreover, their limits are related to each other via the Marčenko–Pastur equation (1.3). Does this property also hold for  $\Sigma_p^{\text{RCV}}$ ? Our first result (Proposition 3) shows that even in the most ideal case when the covolatility process has the form  $\Theta_t = \gamma_t \cdot I_{p \times p}$  for some deterministic (scalar) function  $\gamma_t$ , such convergence results may *not* hold for  $\Sigma_p^{\text{RCV}}$ . In particular, the limit of  $F^{\Sigma_p^{\text{RCV}}}$  (when it exists) changes according to how the covolatility process evolves over time.

This leads to the following natural and interesting question: how does the LSD of RCV matrix depend on the time-variability of the covolatility process? Answering this question in a general context without putting any structural assumption on the covolatility process seems to be rather challenging, if not impossible. For a class  $\mathcal{C}$  (see Section 2) of processes, we do establish a result for RCV matrices that’s analogous to the Marčenko–Pastur theorem (see Proposition 5), which demonstrates clearly how the time-variability of the covolatility process affects the LSD of RCV matrix. Proposition 5 is proved based on Theorem 1, which is a Marčenko–Pastur type theorem for *weighted* sample covariance matrices. These results, in principle, allow one to recover the LSD of ICV matrix based on that of RCV matrix.

Estimating high dimensional ICV matrices based on high frequency data has only recently started to gain attention. See, for example, Wang and Zou (2010); Tao et al. (2011) who made use of data over long time horizons by proposing a method incorporating low-frequency dynamics; and Fan, Li and Yu (2011) who studied the estimation of ICV matrices for portfolio allocation under gross exposure constraint. In Wang and Zou (2010), under sparsity assumptions on the ICV matrix, banding/thresholding was innovatively used to construct consistent estimators of the ICV matrix in the spectral norm sense. In particular, when the sparsity assumptions are satisfied, their estimators share the same LSD as the ICV matrix. It remains an open question that when the sparsity assumptions are not satisfied, whether one can still make good inference about the spectrum of ICV matrix. For processes

in class  $\mathcal{C}$  (see Section 2), whose ICV matrices do not need to be sparse, we propose a new estimator, the *time-variation adjusted realized covariance* (TVARCV) matrix. We show that the TVARCV matrix has the desirable property that its LSD exists provided that the LSD of ICV matrix exists, and furthermore, the two LSDs are related to each other via the Marčenko–Pastur equation (1.3) (see Theorem 2). Therefore, the TVARCV matrix can be used, for example, to recover the LSD of ICV matrix by inverting the Marčenko–Pastur equation using existing algorithms.

The rest of the paper is organized as the following: theoretical results are presented in Section 2, proofs are given in Section 3, simulation studies in Section 4, and conclusion and discussions in Section 5.

*Notation.* For any matrix  $A$ ,  $\|A\| = \sqrt{\lambda_{\max}(AA^*)}$  denotes its spectral norm. For any Hermitian matrix  $A$ ,  $F^A$  stands for its ESD. For two matrices  $A$  and  $B$ , we write  $A \leq B$  ( $A \geq B$ , resp.) if  $B - A$  ( $A - B$ , resp.) is a nonnegative definite matrix. For any interval  $I \subseteq [0, \infty)$ , and any metric space  $S$ ,  $D(I; S)$  stands for the space of càdlàg functions from  $I$  to  $S$ . Additionally,  $i = \sqrt{-1}$  stands for the imaginary unit, and for any  $z \in \mathbb{C}$ , we write  $\operatorname{Re}(z)$ ,  $\operatorname{Im}(z)$  as its real part and imaginary part, respectively, and  $\bar{z}$  as its complex conjugate. We also denote  $\mathbb{R}_+ = \{a \in \mathbb{R} : a > 0\}$ ,  $\mathbb{C}_+ = \{z \in \mathbb{C} : \operatorname{Re}(z) > 0\}$  and  $Q_1 = \{z \in \mathbb{C} : \operatorname{Re}(z) \geq 0, \operatorname{Im}(z) \geq 0\}$ . We follow the custom of writing  $f \sim g$  to mean that the ratio  $f/g$  converges to 1. Finally, throughout the paper,  $c, C, C_1, C'$  etc. denote generic constants whose values may change from line to line.

## 2. Main results.

**2.1. Dependence of the LSD of RCV matrix on the time-variability of co-volatility process.** Proposition 1 asserts that the ESD of sample covariance matrix converges to a limiting distribution which is uniquely determined by the LSD of the underlying covariance matrix. Unfortunately, Proposition 1 does not apply to our case, since the observations  $\Delta \mathbf{X}_\ell$  under our general diffusion process setting are not i.i.d. Proposition 3 below shows that even in the following most ideal case, the RCV matrix does not have the desired convergence property.

**PROPOSITION 3.** *Suppose that for all  $p$ ,  $\mathbf{X}_t = \mathbf{X}_t^{(p)}$  is a  $p$ -dimensional process satisfying*

$$(2.1) \quad d\mathbf{X}_t = \gamma_t d\mathbf{W}_t, \quad t \in [0, 1],$$

where  $\gamma_t > 0$  is a nonrandom (scalar) càdlàg process. Let  $\sigma^2 = \int_0^1 \gamma_t^2 dt$ , and so that the ICV matrix  $\Sigma_p$  is  $\sigma^2 I_{p \times p}$ . Assume further that the observation times  $\tau_{n,\ell}$  are equally spaced, that is,  $\tau_{n,\ell} = \ell/n$ , and that the RCV ma-

trix  $\Sigma_p^{\text{RCV}}$  is defined by (1.2). Then so long as  $\gamma_t$  is not constant on  $[0, 1]$ , for any  $\varepsilon > 0$ , there exists  $y_c = y_c(\gamma, \varepsilon) > 0$  such that if  $\lim p/n = y \geq y_c$ ,

$$(2.2) \quad \limsup F^{\Sigma_p^{\text{RCV}}}(b(y) + \sigma^2 \varepsilon) < 1 \quad \text{almost surely.}$$

In particular,  $F^{\Sigma_p^{\text{RCV}}}$  does not converge to the Marčenko–Pastur law  $\text{MP}^{(y, \sigma^2)}$ .

Observe that  $\text{MP}^{(y, \sigma^2)}$  is the LSD of RCV matrix when  $\gamma_t \equiv \sigma$ . The main message of Proposition 3 is that, the LSD of RCV matrix depends on the whole covolatility process *not only through*  $\Sigma_p$ , *but also on how the covolatility process varies in time*. It will also be clear from the proof of Proposition 3 (Section 3.2) that, the more “volatile” the covolatility process is, the further away the LSD is from the Marčenko–Pastur law  $\text{MP}^{(y, \sigma^2)}$ . This is also illustrated in the simulation study in Section 4.

**2.2. The class  $\mathcal{C}$ .** To understand the behavior of the ESD of RCV matrix more clearly, we next focus on a special class of diffusion processes for which we (i) establish a Marčenko–Pastur type theorem for RCV matrices; and (ii) propose an alternative estimator of ICV matrix.

**DEFINITION 1.** Suppose that  $\mathbf{X}_t$  is a  $p$ -dimensional process satisfying (1.1), and  $\Theta_t$  is càdlàg. We say that  $\mathbf{X}_t$  belongs to class  $\mathcal{C}$  if, almost surely, there exist  $(\gamma_t) \in D([0, 1]; \mathbb{R})$  and  $\Lambda$  a  $p \times p$  matrix satisfying  $\text{tr}(\Lambda \Lambda^T) = p$  such that

$$(2.3) \quad \Theta_t = \gamma_t \Lambda.$$

Observe that if (2.3) holds, then the ICV matrix  $\Sigma_p = \int_0^1 \gamma_t^2 dt \cdot \Lambda \Lambda^T$ . We note that  $\Lambda$  does not need to be sparse, hence neither does  $\Sigma_p$ .

A special case is when  $\Lambda = I_{p \times p}$ . This type of process is studied in Proposition 3 and in the simulation studies in Section 4.

A more interesting case is the following.

**PROPOSITION 4.** Suppose that  $X_t^{(j)}$  satisfy

$$(2.4) \quad dX_t^{(j)} = \mu_t^{(j)} dt + \sigma_t^{(j)} dW_t^{(j)}, \quad j = 1, \dots, p,$$

where  $\mu_t^{(j)}, \sigma_t^{(j)} \in D([0, 1]; \mathbb{R})$  are the drift and volatility processes for stock  $j$ , and  $W_t^{(j)}$ ’s are (one-dimensional) standard Brownian motions. If the following conditions hold:

(i) the correlation matrix process of  $(W_t^{(j)})$

$$(2.5) \quad R_t := \left( \frac{\langle W^{(j)}, W^{(k)} \rangle_t}{t} \right)_{1 \leq j, k \leq p} := (r^{(jk)})_{1 \leq j, k \leq p}$$

is constant in  $t \in (0, 1]$ ;

- (ii)  $r^{(jk)} \neq 0$  for all  $1 \leq j, k \leq p$ ; and
- (iii) the correlation matrix process of  $(X_t^{(j)})$

$$(2.6) \quad \left( \frac{\int_0^t \sigma_s^{(j)} \sigma_s^{(k)} d\langle W^{(j)}, W^{(k)} \rangle_s}{\sqrt{\int_0^t (\sigma_s^{(j)})^2 ds} \cdot \sqrt{\int_0^t (\sigma_s^{(k)})^2 ds}} \right)_{1 \leq j, k \leq p} := (\rho^{(jk)})_{1 \leq j, k \leq p}$$

is constant in  $t \in (0, 1]$ ;

then  $(X_t^{(j)})$  belongs to class  $\mathcal{C}$ .

The proof is given in the supplementary article [Zheng and Li (2011)].

Equation (2.4) is another common way of representing multi-dimensional log-price processes. We note that if  $X_t^{(j)}$  are log price processes, then over short time period, say, one day, it is reasonable to assume that the correlation structure of  $(X_t^{(j)})$  does not change, hence by this proposition,  $(X_t^{(j)})$  belongs to class  $\mathcal{C}$ .

Observe that if a diffusion process  $\mathbf{X}_t$  belongs to class  $\mathcal{C}$ , the drift process  $\boldsymbol{\mu}_t \equiv 0$ , and  $\tau_{n,\ell}$ 's and  $\gamma_t$  are independent of  $\mathbf{W}_t$ , then

$$\Delta \mathbf{X}_\ell = \int_{\tau_{n,\ell-1}}^{\tau_{n,\ell}} \gamma_t \Lambda d\mathbf{W}_t \stackrel{d}{=} \sqrt{\int_{\tau_{n,\ell-1}}^{\tau_{n,\ell}} \gamma_t^2 dt} \cdot \check{\Sigma}^{1/2} \cdot \mathbf{Z}_\ell,$$

where “ $\stackrel{d}{=}$ ” stands for “equal in distribution,”  $\check{\Sigma}^{1/2}$  is the nonnegative square root matrix of  $\check{\Sigma} := \Lambda \Lambda^T$ , and  $\mathbf{Z}_\ell = (Z_\ell^{(1)}, \dots, Z_\ell^{(p)})^T$  consists of independent standard normals. Therefore the RCV matrix

$$\Sigma_p^{\text{RCV}} = \sum_{\ell=1}^n \Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T \stackrel{d}{=} \sum_{\ell=1}^n w_\ell^n \cdot \check{\Sigma}^{1/2} \mathbf{Z}_\ell (\mathbf{Z}_\ell)^T \check{\Sigma}^{1/2},$$

where  $w_\ell^n = \int_{\tau_{n,\ell-1}}^{\tau_{n,\ell}} \gamma_t^2 dt$ . This is similar to the  $S_p$  in Proposition 1, except that here the “weights”  $w_\ell^n$  may vary in  $\ell$ , while in Proposition 1 the “weights” are constantly  $1/n$ . Motivated by this observation we develop the following Marčenko–Pastur type theorems for weighted sample covariance matrices and RCV matrices.

### 2.3. Marčenko–Pastur type theorems for weighted sample covariance matrices and RCV matrices.

**THEOREM 1.** *Suppose that assumptions (ii) and (iv) in Proposition 1 hold. Assume further that:*

(A.i') for  $p = 1, 2, \dots$  and  $1 \leq \ell \leq n$ ,  $\mathbf{Z}_\ell^{(p)} = (Z_\ell^{(p,j)})_{1 \leq j \leq p}$  with  $Z_\ell^{(p,j)}$  i.i.d. with mean 0, variance 1 and finite moments of all orders;



(A.iii')  $\Sigma_p$  is a (possibly random) nonnegative definite  $p \times p$  matrix such that its ESD  $F^{\Sigma_p}$  converges almost surely in distribution to a probability distribution  $H$  on  $[0, \infty)$  as  $p \rightarrow \infty$ ; moreover,  $H$  has a finite second moment;

(A.v) the weights  $w_\ell^n, 1 \leq \ell \leq n, n = 1, 2, \dots$ , are all positive, and there exists  $\kappa < \infty$  such that the rescaled weights  $(nw_\ell^n)$  satisfy

$$\max_n \max_{\ell=1, \dots, n} (nw_\ell^n) \leq \kappa;$$

moreover, almost surely, there exists a process  $w_s \in D([0, 1]; \mathbb{R}_+)$  such that

$$(2.7) \quad \lim_n \sum_{1 \leq \ell \leq n} \int_{(\ell-1)/n}^{\ell/n} |nw_\ell^n - w_s| ds = 0;$$

(A.vi) there exists a sequence  $\eta_p = o(p)$  and a sequence of index sets  $\mathcal{I}_p$  satisfying  $\mathcal{I}_p \subset \{1, \dots, p\}$  and  $\#\mathcal{I}_p \leq \eta_p$  such that for all  $n$  and all  $\ell$ ,  $w_\ell^n$  may depend on  $\mathbf{Z}_\ell^{(p)}$  but only on  $\{Z_\ell^{(p,j)} : j \in \mathcal{I}_p\}$ ;

(A.vii) there exist  $C < \infty$  and  $\delta < 1/6$  such that for all  $p$ ,  $\|\Sigma_p\| \leq Cp^\delta$  almost surely.

Define  $S_p = \sum_{\ell=1}^n w_\ell^n \cdot \Sigma_p^{1/2} \mathbf{Z}_\ell^{(p)} (\mathbf{Z}_\ell^{(p)})^T \Sigma_p^{1/2}$ . Then, almost surely, the ESD of  $S_p$  converges in distribution to a probability distribution  $F^w$ , which is determined by  $H$  and  $(w_s)$  in that its Stieltjes transform  $m_{F^w}(z)$  is given by

$$(2.8) \quad m_{F^w}(z) = -\frac{1}{z} \int_{\tau \in \mathbb{R}} \frac{1}{\tau M(z) + 1} dH(\tau),$$

where  $M(z)$ , together with another function  $\tilde{m}(z)$ , uniquely solve the following equation in  $\mathbb{C}_+ \times \mathbb{C}_+$ :

$$(2.9) \quad \begin{cases} M(z) = -\frac{1}{z} \int_0^1 \frac{w_s}{1 + y\tilde{m}(z)w_s} ds, \\ \tilde{m}(z) = -\frac{1}{z} \int_{\tau \in \mathbb{R}} \frac{\tau}{\tau M(z) + 1} dH(\tau). \end{cases}$$

REMARK 2. Assumption (A.i') can undoubtedly be weakened, for example, by using the truncation and centralization technique as in Silverstein and Bai (1995) and Silverstein (1995); or, a closer look at the proof of Theorem 1 indicates that as long as  $Z_\ell^{(p,j)}$  has finite moments up to order  $k > 6/(1 - 6\delta)$ , the theorem is true and can be proved by exactly the same argument.

REMARK 3. If  $w_\ell^n \equiv 1/n$ , then  $w_s \equiv 1$ , and Theorem 1 reduces to Proposition 1. Moreover, if  $w_s$  is not constant, that is,  $w_s \not\equiv \int_0^1 w_t dt$  on  $[0, 1]$ , then except in the trivial case when  $H$  is a delta measure at 0, the LSD  $F^w \neq F$ , where  $F$  is the LSD in Proposition 1 determined by  $H(\cdot / \int_0^1 w_t dt)$ . See the supplementary article [Zheng and Li (2011)] for more details.

Theorem 1 is proved in Section 3.3.

A direct consequence of this theorem and Lemma 1 below is the following Marčenko–Pastur type result for RCV matrices for diffusion processes in class  $\mathcal{C}$ . We note that, thanks to Lemma 1 below (see the remark after the proof of Lemma 1 for more explanations), regarding the drift process, except requiring them to be uniformly bounded, we put no additional assumption on them: they can be, for example, stochastic, càdlàg and dependant with each other. Furthermore, we allow for dependence between the covolatility process and the underlying Brownian motion—in other words, we allow for the leverage effect. In the special case when  $\gamma_t^{(p)}$  does not change in  $p$ , is nonrandom and bounded, and the observation times are equally spaced, the (rather technical) assumptions (B.iii) and (B.iv) below are trivially satisfied.

**PROPOSITION 5.** *Suppose that for all  $p$ ,  $\mathbf{X}_t^{(p)}$  is a  $p$ -dimensional process in class  $\mathcal{C}$  for some drift process  $\mu_t^{(p)} = (\mu_t^{(p,1)}, \dots, \mu_t^{(p,p)})^T$ , covolatility process  $\Theta_t^{(p)} = \gamma_t^{(p)} \Lambda^{(p)}$  and  $p$ -dimensional Brownian motion  $\mathbf{W}_t^{(p)} = (W_t^{(p,1)}, \dots, W_t^{(p,p)})^T$ . Suppose further that:*

(B.i) *there exists  $C_0 < \infty$  such that for all  $p$  and all  $j = 1, \dots, p$ ,  $|\mu_t^{(p,j)}| \leq C_0$  for all  $t \in [0, 1]$  almost surely;*

(B.ii)  *$\check{\Sigma}_p = \Lambda^{(p)}(\Lambda^{(p)})^T$  satisfies assumption (A.iii') and (A.vii) in Theorem 1;*

(B.iii) *there exists a sequence  $\eta_p = o(p)$  and a sequence of index sets  $\mathcal{I}_p$  satisfying  $\mathcal{I}_p \subset \{1, \dots, p\}$  and  $\#\mathcal{I}_p \leq \eta_p$  such that  $\gamma_t^{(p)}$  may depend on  $\mathbf{W}_t^{(p)}$  but only on  $\{W_t^{(p,j)} : j \in \mathcal{I}_p\}$ ; moreover, there exists  $C_1 < \infty$  such that for all  $p$ ,  $|\gamma_t^{(p)}| \leq C_1$  for all  $t \in [0, 1]$  almost surely; additionally, almost surely, there exists  $(\gamma_t) \in D([0, 1]; \mathbb{R})$  such that*

$$\lim_p \int_0^1 |\gamma_t^{(p)} - \gamma_t| dt = 0;$$

(B.iv) *the observation times  $\tau_{n,\ell}$  are independent of  $\mathbf{X}_t$ ; moreover, there exists  $\kappa < \infty$  such that the observation durations  $\Delta\tau_{n,\ell} := \tau_{n,\ell} - \tau_{n,\ell-1}$  satisfy*

$$\max_n \max_{\ell=1, \dots, n} (n \cdot \Delta\tau_{n,\ell}) \leq \kappa;$$

*additionally, almost surely, there exists a process  $v_s \in C([0, 1]; \mathbb{R}_+)$  such that*

$$\tau_{n, [ns]} \rightarrow \Upsilon_s := \int_0^s v_r dr \quad \text{as } n \rightarrow \infty \text{ for all } 0 \leq s \leq 1,$$

*where for any  $x$ ,  $[x]$  stands for its integer part.*

*Then, as  $p \rightarrow \infty$ ,  $F^{\Sigma_p^{\text{RCV}}}$  converges almost surely to a probability distribution  $F^w$  as specified in Theorem 1 for  $w_s = (\gamma_{\Upsilon_s})^2 v_s$ .*

Proposition 5 demonstrates explicitly how the LSD of RCV matrix depends on the time-variability of the covolatility process. Hence, the RCV matrix by itself cannot be used to make robust inference for the ESD  $F^{\Sigma_p}$  of the ICV matrix. If  $(\gamma_s)$  [and hence  $w_s = (\gamma_s)^2 v_s$ ] is known, then in principle, the equations (2.8) and (2.9) can be used to recover  $F^{\Sigma_p}$ . However, in general,  $(\gamma_s)$  is unknown and estimating the process  $(\gamma_s)$  can be challenging and will bring in more complication in the inference. Moreover, the equations (2.8) and (2.9) are different from and more complicated than the classical Marčenko–Pastur equation (1.3), and in order to recover  $F^{\Sigma_p}$  based on these equations, one has to extend existing algorithms [El Karoui (2008), Mestre (2008) and Bai, Chen and Yao (2010) etc.] which are designed for (1.3). Developing such an algorithm is of course of great interest, but we shall not pursue this in the present article. We shall instead propose an alternative estimator which overcomes these difficulties.

2.4. *Time-variation adjusted realized covariance (TVARCV) matrix.* Suppose that a diffusion process  $\mathbf{X}_t$  belongs to class  $\mathcal{C}$ . We define the *time-variation adjusted realized covariance* (TVARCV) matrix as follows:

$$(2.10) \quad \hat{\Sigma}_p := \frac{\text{tr}(\Sigma_p^{\text{RCV}})}{n} \cdot \sum_{\ell=1}^n \frac{\Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T}{|\Delta \mathbf{X}_\ell|^2} = \frac{\text{tr}(\Sigma_p^{\text{RCV}})}{p} \tilde{\Sigma}_p,$$

where for any vector  $\mathbf{v}$ ,  $|\mathbf{v}|$  stands for its Euclidean norm, and

$$(2.11) \quad \tilde{\Sigma}_p := \frac{p}{n} \cdot \sum_{\ell=1}^n \frac{\Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T}{|\Delta \mathbf{X}_\ell|^2}.$$

Let us first explain  $\tilde{\Sigma}_p$ . Consider the simplest case when  $\mu_t \equiv 0$ ,  $\gamma_t$  deterministic,  $\Lambda_t \equiv I_{p \times p}$ , and  $\tau_{n,\ell} = \ell/n$ ,  $\ell = 0, 1, \dots, n$ . In this case,  $\Delta \mathbf{X}_\ell = \sqrt{\int_{(\ell-1)/n}^{\ell/n} \gamma_t^2 dt} \cdot \mathbf{Z}_\ell / \sqrt{n}$  where  $\mathbf{Z}_\ell = (Z_\ell^{(1)}, \dots, Z_\ell^{(p)})^T$  and  $Z_\ell^{(j)}$ 's are i.i.d. standard normal. Hence,  $\Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T / |\Delta \mathbf{X}_\ell|^2 = \mathbf{Z}_\ell (\mathbf{Z}_\ell)^T / |\mathbf{Z}_\ell|^2$ . However, as  $p \rightarrow \infty$ ,  $|\mathbf{Z}_\ell|^2 \sim p$ , hence  $\tilde{\Sigma}_p \sim 1/n \cdot \sum_{\ell=1}^n \mathbf{Z}_\ell (\mathbf{Z}_\ell)^T$ , the latter being the usual sample covariance matrix. We will show that, first,  $\text{tr}(\Sigma_p^{\text{RCV}}) \sim \text{tr}(\Sigma_p)$ ; and second, if  $\mathbf{X}_t$  belongs to class  $\mathcal{C}$  and satisfies certain additional assumptions, then the LSD of  $\hat{\Sigma}_p$  is related to that of  $\check{\Sigma}_p$  via the Marčenko–Pastur equation (1.3), where

$$(2.12) \quad \check{\Sigma}_p = \frac{p}{\text{tr}(\Sigma_p)} \Sigma_p = \Lambda \Lambda^T.$$

Hence, the LSD of  $\hat{\Sigma}_p$  is also related to that of  $\Sigma_p$  via the same Marčenko–Pastur equation.

We now state our assumptions. Observe that about the drift process, again, except requiring them to be uniformly bounded, we put no additional assumption. Furthermore, we allow for the dependence between the

covolatility process and the underlying Brownian motion, namely, the leverage effect.

*Assumptions:*

(C.i) there exists  $C_0 < \infty$  such that for all  $p$  and all  $j = 1, \dots, p$ ,  $|\mu_t^{(p,j)}| \leq C_0$  for all  $t \in [0, 1]$  almost surely;

(C.ii) there exist constants  $C_1 < \infty, 0 \leq \delta_1 < 1/2$ , a sequence  $\eta_p < C_1 p^{\delta_1}$  and a sequence of index sets  $\mathcal{I}_p$  satisfying  $\mathcal{I}_p \subset \{1, \dots, p\}$  and  $\#\mathcal{I}_p \leq \eta_p$  such that  $\gamma_t^{(p)}$  may depend on  $\mathbf{W}_t^{(p)}$  but only on  $\{W_t^{(p,j)} : j \in \mathcal{I}_p\}$ ; moreover, there exists  $C_2 < \infty$  such that for all  $p$ ,  $|\gamma_t^{(p)}| \in (1/C_2, C_2)$  for all  $t \in [0, 1]$  almost surely;

(C.iii) there exists  $C_3 < \infty$  such that for all  $p$  and for all  $j$ , the individual volatilities  $\sigma_t^{(j)} = \sqrt{(\gamma_t^{(p)})^2 \cdot \sum_{k=1}^p (\Lambda_{jk}^{(p)})^2} \in (1/C_3, C_3)$  for all  $t \in [0, 1]$  almost surely;

(C.iv)  $\lim_{p \rightarrow \infty} \text{tr}(\Sigma_p)/p (= \lim_{p \rightarrow \infty} \int_0^1 (\gamma_t^{(p)})^2 dt) := \theta > 0$  almost surely;

(C.v) almost surely, as  $p \rightarrow \infty$ , the ESD  $F^{\Sigma_p}$  converges to a probability distribution  $H$  on  $[0, \infty)$ ;

(C.vi) there exist  $C_5 < \infty$  and  $0 \leq \delta_2 < 1/2$  such that for all  $p$ ,  $\|\Sigma_p\| \leq C_5 p^{\delta_2}$  almost surely;

(C.vii) the  $\delta_1$  in (C.ii) and  $\delta_2$  in (C.vi) satisfy that  $\delta_1 + \delta_2 < 1/2$ ;

(C.viii)  $p/n \rightarrow y \in (0, \infty)$  as  $p \rightarrow \infty$ ; and

(C.ix) there exists  $C_4 < \infty$  such that for all  $n$ ,

$$\max_{1 \leq \ell \leq n} n \cdot (\tau_{n,\ell} - \tau_{n,\ell-1}) \leq C_4 \quad \text{almost surely;}$$

moreover,  $\tau_{n,\ell}$ 's are independent of  $\mathbf{X}_t$ .

We have the following convergence theorem regarding the ESD of our proposed estimator TVARCV matrix  $\hat{\Sigma}_p$ .

**THEOREM 2.** *Suppose that for all  $p$ ,  $\mathbf{X}_t = \mathbf{X}_t^{(p)}$  is a  $p$ -dimensional process in class  $\mathcal{C}$  for some drift process  $\mu_t^{(p)} = (\mu_t^{(p,1)}, \dots, \mu_t^{(p,p)})^T$ , covolatility process  $\Theta_t^{(p)} = \gamma_t^{(p)} \Lambda^{(p)}$  and  $p$ -dimensional Brownian motion  $\mathbf{W}_t^{(p)}$ , which satisfy assumptions (C.i)~(C.vii) above. Suppose also that  $p$  and  $n$  satisfy (C.viii), and the observation times satisfy (C.ix). Let  $\hat{\Sigma}_p$  be as in (2.10). Then, as  $p \rightarrow \infty$ ,  $F^{\hat{\Sigma}_p}$  converges almost surely to a probability distribution  $F$ , which is determined by  $H$  through Stieltjes transforms via the same Marčenko–Pastur equation (1.3) as in Proposition 1.*

The proof of Theorem 2 is given in Section 3.4.

The LSD  $H$  of the targeting ICV matrix is in general not the same as the LSD  $F$ , but can be recovered from  $F$  based on equation (1.3). In practice,

when one has only finite number of samples, the articles [El Karoui (2008), Mestre (2008) and Bai, Chen and Yao (2010) etc.] studied the estimation of the population spectral distribution based on the sample covariance matrices. In particular, applying Theorem 2 of El Karoui (2008) to our case yields.

**COROLLARY 1.** *Let  $H_p = F^{\widehat{\Sigma}_p}$ , and define  $\widehat{H}_p$  as in Theorem 2 of El Karoui (2008). If  $\|\Sigma_p\|$  are bounded in  $p$ , then, as  $p \rightarrow \infty$ ,  $\widehat{H}_p \rightarrow H$  almost surely.*

Therefore, when the dimension  $p$  is large, based on the ESD of TVARCV matrix  $\widehat{\Sigma}_p$ , we can estimate the spectrum of underlying ICV matrix  $\Sigma_p$  well.

### 3. Proofs.

**3.1. Preliminaries.** We collect some either elementary or well-known facts in the following. The proofs are given in the supplemental article [Zheng and Li (2011)].

**LEMMA 1.** *Suppose that for each  $p$ ,  $\mathbf{v}_\ell^{(p)} = (v_\ell^{(p,1)}, \dots, v_\ell^{(p,p)})^T$  and  $\mathbf{w}_\ell^{(p)} = (w_\ell^{(p,1)}, \dots, w_\ell^{(p,p)})$ ,  $\ell = 1, \dots, n$ , are all  $p$ -dimensional vectors. Define*

$$\widetilde{S}_n = \sum_{\ell=1}^n (\mathbf{v}_\ell^{(p)} + \mathbf{w}_\ell^{(p)}) \cdot (\mathbf{v}_\ell^{(p)} + \mathbf{w}_\ell^{(p)})^T \quad \text{and} \quad S_n = \sum_{\ell=1}^n \mathbf{w}_\ell^{(p)} (\mathbf{w}_\ell^{(p)})^T.$$

*If the following conditions are satisfied:*

- (i)  $n = n(p)$  with  $\lim_{p \rightarrow \infty} p/n = y > 0$ ;
- (ii) *there exists a sequence  $\varepsilon_p = o(1/\sqrt{p})$  such that for all  $p$  and all  $\ell$ , all the entries of  $\mathbf{v}_\ell^{(p)}$  are bounded by  $\varepsilon_p$  in absolute value;*
- (iii)  $\limsup_{p \rightarrow \infty} \text{tr}(S_n)/p < \infty$  almost surely.

*Then  $L(F^{\widetilde{S}_n}, F^{S_n}) \rightarrow 0$  almost surely, where for any two probability distribution functions  $F$  and  $G$ ,  $L(F, G)$  denotes the Levy distance between them.*

**LEMMA 2** [Lemma 2.6 of Silverstein and Bai (1995)]. *Let  $z \in \mathbb{C}$  with  $v = \text{Im}(z) > 0$ ,  $A$  and  $B$  be  $p \times p$  with  $B$  Hermitian, and  $\mathbf{q} \in \mathbb{C}^p$ . Then*

$$|\text{tr}(((B - zI)^{-1} - (B + \tau \mathbf{q} \mathbf{q}^* - zI)^{-1}) \cdot A)| \leq \|A\|/v \quad \text{for all } \tau \in \mathbb{R}.$$

The following two lemmas are similar to Lemma 2.3 in Silverstein (1995).

**LEMMA 3.** *Let  $w \in \mathbb{C}$  with  $\text{Re}(w) \geq 0$ , and  $A$  be an Hermitian nonnegative definite matrix. Then  $\|(wA + I)^{-1}\| \leq 1$ .*

**LEMMA 4.** *Let  $w_1, w_2 \in \mathbb{C}$  with  $\text{Re}(w_1) \geq 0$  and  $\text{Re}(w_2) \geq 0$ ,  $A$  be a  $p \times p$  Hermitian nonnegative definite matrix,  $B$  any  $p \times p$  matrix, and  $\mathbf{q} \in \mathbb{C}^p$ .*

Then:

- (i)  $|\operatorname{tr}(B((w_1A + I)^{-1} - (w_2A + I)^{-1}))| \leq p \cdot |w_1 - w_2| \cdot \|B\| \cdot \|A\|;$
- (ii)  $|\mathbf{q}^*B(w_1A + I)^{-1}\mathbf{q} - \mathbf{q}^*B(w_2A + I)^{-1}\mathbf{q}| \leq |w_1 - w_2| \cdot |\mathbf{q}|^2 \|B\| \cdot \|A\|.$

LEMMA 5. *For any Hermitian matrix  $A$  and  $z \in \mathbb{C}$  with  $\operatorname{Im}(z) = v > 0$ ,  $\|(A - zI)^{-1}\| \leq 1/v$ .*

Both Lemmas 3 and 4 require the real part of  $w$  (or  $w_1, w_2$ ) to be nonnegative. In our proof of Theorem 1, the requirements will be fulfilled thanks to the following lemma.

LEMMA 6. *Let  $z = iv \in \mathbb{C}$  with  $v > 0$ ,  $A$  be a  $p \times p$  Hermitian nonnegative definite matrix,  $\mathbf{q} \in \mathbb{C}^p$ ,  $\tau > 0$ . Then*

$$-\frac{1}{z} \cdot \frac{1}{1 + \tau \mathbf{q}^*(A - zI)^{-1}\mathbf{q}} \in Q_1 = \{z \in \mathbb{C} : \operatorname{Re}(z) \geq 0, \operatorname{Im}(z) \geq 0\}.$$

LEMMA 7. *Let  $z = iv \in \mathbb{C}$  with  $v > 0$ ,  $A$  be any  $p \times p$  matrix, and  $B$  be a  $p \times p$  Hermitian nonnegative definite matrix. Then  $\operatorname{tr}(A(B - zI)^{-1}A^*) \in Q_1$ .*

LEMMA 8. *Suppose that  $w_s \in D([0, 1]; \mathbb{R}_+)$ . Then for any  $y \in \mathbb{C}$ , the equation*

$$\int_0^1 \frac{1}{1 + zw_s} ds = y$$

*admits at most one solution in  $Q_1$ .*

The following result is an immediate consequence of Lemma 2.7 of Bai and Silverstein (1998).

LEMMA 9. *For  $\mathbf{X} = (X^{(1)}, \dots, X^{(p)})^T$  where  $X^{(j)}$ 's are i.i.d. random variables such that  $EX^{(1)} = 0$ ,  $E|X^{(1)}|^2 = 1$ , and  $E|X^{(1)}|^{2k} < \infty$  for some  $2 \leq k \in \mathbb{N}$ , there exists  $C_k \geq 0$ , depending only on  $k$ ,  $E|X^{(1)}|^4$  and  $E|X^{(1)}|^{2k}$ , such that for any  $p \times p$  nonrandom matrix  $A$ ,*

$$E|\mathbf{X}^*A\mathbf{X} - \operatorname{tr}(A)|^{2k} \leq C_k(\operatorname{tr}(AA^*))^k \leq C_k p^k \|A\|^{2k}.$$

PROPOSITION 6 [Theorem 2 of Geronimo and Hill (2003)]. *Suppose that  $P_n$  are real probability measures with Stieltjes transforms  $m_n(z)$ . Let  $K \subseteq \mathbb{C}_+$  be an infinite set with a limit point in  $\mathbb{C}_+$ . If  $\lim m_n(z) := m(z)$  exists for all  $z \in K$ , then there exists a probability measure  $P$  with Stieltjes transform  $m(z)$  if and only if*

$$(3.1) \quad \lim_{v \rightarrow \infty} iv \cdot m(iv) = -1,$$

*in which case  $P_n \rightarrow P$  in distribution.*

3.2. *Proof of Proposition 3.* By assumption,  $\gamma_t$  is positive and non-constant on  $[0, 1]$ , and is càdlàg, in particular, right-continuous; moreover,  $\int_0^1 \gamma_t^2 dt = \sigma^2$ . Hence, there exists  $\delta > 0$  and  $[c, d] \subseteq [0, 1]$  such that

$$\gamma_t \geq \sigma(1 + \delta) \quad \text{for all } t \in [c, d].$$

Therefore, if  $[(\ell - 1)/n, \ell/n] \subseteq [c, d]$ ,

$$\Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T \stackrel{d}{=} \int_{(\ell-1)/n}^{\ell/n} \gamma_t^2 dt \cdot \mathbf{Z}_\ell (\mathbf{Z}_\ell)^T \geq \frac{(1 + \delta)^2}{n} \cdot \sigma^2 \mathbf{Z}_\ell (\mathbf{Z}_\ell)^T,$$

where  $\mathbf{Z}_\ell = (Z_\ell^{(1)}, \dots, Z_\ell^{(p)})^T$  consists of independent standard normals. Hence, if we let  $J_n = \{\ell : [(\ell - 1)/n, \ell/n] \subseteq [c, d]\}$  and

$$\Gamma_p = \sum_{\ell \in J_n} \Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T, \quad \Lambda_p = \frac{\sigma^2}{(n(d - c))} \cdot \sum_{\ell \in J_n} \mathbf{Z}_\ell (\mathbf{Z}_\ell)^T,$$

then for any  $x \geq 0$ , by Weyl's Monotonicity theorem [see, e.g., Corollary 4.3.3 in Horn and Johnson (1990)],

$$(3.2) \quad F^{\Sigma_p^{\text{RCV}}}(x) \leq F^{\Gamma_p}(x) \leq F^{\Lambda_p}(x / [(1 + \delta)^2 (d - c)]).$$

Now note that  $\#J_n \sim (d - c)n$ , hence if  $p/n \rightarrow y$ , by Proposition 2,  $F^{\Lambda_p}$  will converge almost surely to the Marčenko–Pastur law with ratio index  $y' = y/(d - c)$  and scale index  $\sigma^2$ , which has density on  $[a(y'), b(y')]$  with functions  $a(\cdot)$  and  $b(\cdot)$  defined by (1.4). By the formula of  $b(\cdot)$ ,

$$(1 + \delta)^2 (d - c) b(y') = (1 + \delta) \cdot \sigma^2 (1 + \delta) (y + 2\sqrt{(d - c)y} + d - c).$$

Hence, for any  $\varepsilon > 0$ , there exists  $y_c > 0$  such that for all  $y \geq y_c$ ,

$$(1 + \delta)^2 (d - c) b(y') \geq (1 + \delta) \cdot \sigma^2 ((1 + \sqrt{y})^2 + \varepsilon) = (1 + \delta) (b(y) + \sigma^2 \varepsilon),$$

that is,

$$\frac{(b(y) + \sigma^2 \varepsilon)}{(1 + \delta)^2 (d - c)} \leq \frac{b(y')}{1 + \delta}.$$

By (3.2), when the above inequality holds,

$$\limsup F^{\Sigma_p^{\text{RCV}}}(b(y) + \sigma^2 \varepsilon) \leq \text{MP}^{(y', \sigma^2)}(b(y') / (1 + \delta)) < 1.$$

3.3. *Proof of Theorem 1.* To prove Theorem 1, following the strategies in Marčenko and Pastur (1967), Silverstein (1995), Silverstein and Bai (1995), we will work with Stieltjes transforms.

PROOF OF THEOREM 1. For notational ease, we shall sometimes omit the sub/superscripts  $p$  and  $n$  in the arguments below: thus, we write  $\mathbf{Z}_\ell$

instead of  $\mathbf{Z}_\ell^{(p)}$ ,  $w_\ell$  instead of  $w_\ell^n$ ,  $\Sigma$  instead of  $\Sigma_p$ ,  $S$  instead of  $S_p$ , etc. Also recall that  $y_n = p/n$ , which converges to  $y > 0$ .

By assumption (A.vi) we may, without loss of generality, assume that the weights  $w_\ell$  are independent of  $\mathbf{Z}_\ell$ 's. This is because, if we let  $\tilde{\mathbf{Z}}_\ell$  be the result of replacing  $Z_\ell^{(p,j)}$ ,  $j \in \mathcal{I}_p$ , with independent random variables with the same distribution that are also independent of  $w_\ell$ , and  $\tilde{S} := \sum_{\ell=1}^n w_\ell \cdot \Sigma^{1/2} \tilde{\mathbf{Z}}_\ell (\tilde{\mathbf{Z}}_\ell)^T \Sigma^{1/2}$ , then  $\text{rank}(\tilde{S} - S) \leq 2\eta_p$ , and so by the rank inequality

$$\|F^A - F^B\| \leq \frac{\text{rank}(A - B)}{p} \quad \text{for any } A, B \text{ } p \times p \text{ symmetric matrices}$$

[see, e.g., Lemma 2.2 in Bai (1999)],  $\tilde{S}$  and  $S$  must have the same LSD.

We proceed according to whether  $H$  is a delta measure at 0 or not. If  $H$  is a delta measure at 0, we claim that  $F^w$  is also a delta measure at 0, and the conclusion of the theorem holds. The reason is as follows. By assumption (A.v),

$$S = \sum_{\ell=1}^n w_\ell \cdot \Sigma^{1/2} \mathbf{Z}_\ell (\mathbf{Z}_\ell)^T \Sigma^{1/2} \leq \frac{\kappa}{n} \sum_{\ell=1}^n \Sigma^{1/2} \mathbf{Z}_\ell (\mathbf{Z}_\ell)^T \Sigma^{1/2} := \kappa \bar{S}.$$

Hence by Weyl's Monotonicity theorem again, for any  $x \geq 0$

$$F^S(x) \geq F^{\bar{S}}(x/\kappa).$$

However, it follows easily from Proposition 1 that  $F^{\bar{S}}$  converges to the delta measure at 0, hence so does  $F^S$ .

Below we assume that  $H$  is *not* a delta measure at 0.

Let  $I = I_{p \times p}$  be the  $p \times p$  identity matrix, and

$$m_n := m_n(z) = \frac{\text{tr}((S - zI)^{-1})}{p}$$

be the Stieltjes transform of  $F^S$ . By Proposition 6, in order to show that  $F^S$  converges, it suffices to prove that for all  $z = iv$  with  $v > 0$  sufficiently large,  $\lim_n m_n(z) := m(z)$  exists, and that  $m(z)$  satisfies condition (3.1).

We first show the convergence of  $m_n(z)$  for  $z = iv$  with  $v > 0$  sufficiently large. Since for all  $n$ ,  $|m_n(z)| \leq 1/v$ , it suffices to show that  $\{m_n(z)\}$  has at most one limit.

For notational ease, we denote by  $\mathbf{r}_\ell = \Sigma^{1/2} \mathbf{Z}_\ell$ . We first show that

$$(3.3) \quad \max_{\ell=1, \dots, n} \|\mathbf{r}_\ell\|^2/p - h = \max_{\ell=1, \dots, n} |\mathbf{Z}_\ell^T \Sigma \mathbf{Z}_\ell/p - h| \rightarrow 0 \quad \text{almost surely,}$$

where  $h = \int_0^\infty x dH(x)$ . In fact, by Lemma 9 and assumptions (A.i') and (A.vii), for any  $k \in \mathbb{N}$ ,

$$E|\mathbf{Z}_\ell^T \Sigma \mathbf{Z}_\ell - \text{tr}(\Sigma)|^{2k} \leq C_k p^k p^{2\delta k} \quad \text{for all } 1 \leq \ell \leq n.$$



Using Markov's inequality we get that for any  $\varepsilon > 0$ ,

$$P(|\mathbf{Z}_\ell^T \Sigma \mathbf{Z}_\ell - \text{tr}(\Sigma)| \geq p\varepsilon) \leq \frac{C_k p^k p^{2\delta k}}{p^{2k} \varepsilon^{2k}} = \frac{C_k \varepsilon^{-2k}}{p^{(1-2\delta)k}} \quad \text{for all } 1 \leq \ell \leq n.$$

Hence, choosing  $k > 2/(1-2\delta)$ , using Borel–Cantelli and that  $n = O(p)$  yield

$$(3.4) \quad \max_{\ell=1, \dots, n} |\mathbf{Z}_\ell^T \Sigma \mathbf{Z}_\ell / p - \text{tr}(\Sigma)/p| \rightarrow 0 \quad \text{almost surely.}$$

The convergence (3.3) follows.

Next, let

$$(3.5) \quad M_n = M_n(z) = -\frac{1}{z} \sum_{\ell=1}^n \frac{w_\ell}{1 + w_\ell \mathbf{r}_\ell^T (S_{(\ell)} - zI)^{-1} \mathbf{r}_\ell},$$

where

$$S_{(\ell)} := \sum_{j \neq \ell} w_j \mathbf{r}_j \mathbf{r}_j^T = S - w_\ell \mathbf{r}_\ell \mathbf{r}_\ell^T.$$

Note that by Lemma 6, for any  $\ell$ ,

$$(3.6) \quad -\frac{1}{z} \frac{1}{1 + w_\ell \mathbf{r}_\ell^T (S_{(\ell)} - zI)^{-1} \mathbf{r}_\ell} \in Q_1.$$

We shall show that

$$(3.7) \quad \frac{1}{p} \text{tr}(-zM_n \Sigma - zI)^{-1} - m_n \rightarrow 0 \quad \text{almost surely.}$$

Observe the following identity: for any  $p \times p$  matrix  $B$ ,  $\mathbf{q} \in \mathbb{R}^p$  and  $\tau \in \mathbb{C}$  for which  $B$  and  $B + \tau \mathbf{q} \mathbf{q}^T$  are both invertible,

$$(3.8) \quad \mathbf{q}^T (B + \tau \mathbf{q} \mathbf{q}^T)^{-1} = \frac{1}{1 + \tau \mathbf{q}^T B^{-1} \mathbf{q}} \mathbf{q}^T B^{-1};$$

see equation (2.2) in Silverstein and Bai (1995). Writing

$$S - zI - (-zM_n \Sigma - zI) = \sum_{\ell=1}^n w_\ell \mathbf{r}_\ell \mathbf{r}_\ell^T - (-zM_n \Sigma),$$

taking the inverse, using (3.8) and the definition (3.5) of  $M_n$  yield

$$\begin{aligned} & (-zM_n \Sigma - zI)^{-1} - (S - zI)^{-1} \\ &= (-zM_n \Sigma - zI)^{-1} \left( \sum_{\ell=1}^n w_\ell \mathbf{r}_\ell \mathbf{r}_\ell^T - (-zM_n \Sigma) \right) (S - zI)^{-1} \\ &= -\frac{1}{z} \sum_{\ell=1}^n \frac{w_\ell}{1 + w_\ell \mathbf{r}_\ell^T (S_{(\ell)} - zI)^{-1} \mathbf{r}_\ell} (M_n \Sigma + I)^{-1} \mathbf{r}_\ell \mathbf{r}_\ell^T (S_{(\ell)} - zI)^{-1} \\ &\quad + \frac{1}{z} \sum_{\ell=1}^n \frac{w_\ell}{1 + w_\ell \mathbf{r}_\ell^T (S_{(\ell)} - zI)^{-1} \mathbf{r}_\ell} (M_n \Sigma + I)^{-1} \Sigma (S - zI)^{-1}. \end{aligned}$$

Taking trace and dividing by  $p$  we get

$$\frac{1}{p} \operatorname{tr}(-zM_n S - zI)^{-1} - m_n = \frac{1}{z} \sum_{\ell=1}^n \frac{w_\ell}{1 + w_\ell \mathbf{r}_\ell^T (S_{(\ell)} - zI)^{-1} \mathbf{r}_\ell} \cdot d_\ell,$$

where

$$d_\ell = \frac{1}{p} (\operatorname{tr}((M_n \Sigma + I)^{-1} \Sigma (S - zI)^{-1}) - \mathbf{r}_\ell^T (S_{(\ell)} - zI)^{-1} (M_n \Sigma + I)^{-1} \mathbf{r}_\ell).$$

By (5.2) in the proof of Lemma 6 in the supplementary article [Zheng and Li (2011)],  $\operatorname{Re}(\mathbf{r}_\ell^T (S_{(\ell)} - zI)^{-1} \mathbf{r}_\ell) \geq 0$ . Hence,

$$(3.9) \quad \left| \frac{1}{1 + w_\ell \mathbf{r}_\ell^T (S_{(\ell)} - zI)^{-1} \mathbf{r}_\ell} \right| \leq 1.$$

Therefore in order to show (3.7), by assumption (A.v), it suffices to prove

$$(3.10) \quad \max_{\ell=1, \dots, n} |d_\ell| \rightarrow 0 \quad \text{almost surely.}$$

Define

$$M_{(\ell)} = M_{(\ell)}(z) = -\frac{1}{z} \sum_{j \neq \ell} \frac{w_j}{1 + w_j \mathbf{r}_j^T (S_{(j, \ell)} - zI)^{-1} \mathbf{r}_j},$$

where  $S_{(j, \ell)} := \sum_{i \neq j, \ell} w_i \mathbf{r}_i \mathbf{r}_i^T = S - (w_j \mathbf{r}_j \mathbf{r}_j^T + w_\ell \mathbf{r}_\ell \mathbf{r}_\ell^T)$ . Observe that for every  $\ell$ ,  $M_{(\ell)}$  is independent of  $\mathbf{Z}_\ell$ .

CLAIM 1. For any  $z = iv$  with  $v > 0$  and any  $\varepsilon < 1/2 - \delta$ ,

$$(3.11) \quad \max_{\ell=1, \dots, n} p^\varepsilon |M_{(\ell)}(z) - M_n(z)| \rightarrow 0 \quad \text{almost surely.}$$

PROOF. Define

$$\tilde{m}_n(z) = \frac{\operatorname{tr}(\Sigma^{1/2} (S - zI)^{-1} \Sigma^{1/2})}{p},$$

which belongs to  $Q_1$  by Lemma 7, and

$$(3.12) \quad \begin{cases} \widetilde{M}_n = \widetilde{M}_n(z) = -\frac{1}{z} \sum_{j=1}^n \frac{w_j}{1 + y_n n w_j \tilde{m}_n(z)}, \\ \widetilde{M}_{(\ell)} = \widetilde{M}_{(\ell)}(z) = -\frac{1}{z} \sum_{j \neq \ell} \frac{w_j}{1 + y_n n w_j \tilde{m}_n(z)} \\ \quad = \widetilde{M}_n(z) + \frac{1}{z} \frac{w_\ell}{1 + y_n n w_\ell \tilde{m}_n(z)}. \end{cases}$$

Then by a similar argument for (3.9) and using assumption (A.v),

$$\max_{\ell=1, \dots, n} |\widetilde{M}_{(\ell)}(z) - \widetilde{M}_n(z)| \leq \frac{\kappa}{nv}.$$

Hence, it suffices to show that

$$(3.13) \quad \begin{aligned} p^\varepsilon |M_n(z) - \widetilde{M}_n(z)| &\rightarrow 0 \quad \text{and} \\ \max_{\ell=1,\dots,n} p^\varepsilon |M_{(\ell)}(z) - \widetilde{M}_{(\ell)}(z)| &\rightarrow 0 \quad \text{almost surely.} \end{aligned}$$

We shall only prove the second convergence. In fact,

$$\begin{aligned} M_{(\ell)}(z) - \widetilde{M}_{(\ell)}(z) &= -\frac{1}{z} \sum_{j \neq \ell} \frac{w_j \cdot y_n n w_j}{(1 + w_j \mathbf{r}_j^T (S_{(j,\ell)} - zI)^{-1} \mathbf{r}_j) \cdot (1 + y_n n w_j \widetilde{m}_n(z))} \cdot \zeta_{j,\ell}, \end{aligned}$$

where

$$\zeta_{j,\ell} = \widetilde{m}_n(z) - \frac{\mathbf{r}_j^T (S_{(j,\ell)} - zI)^{-1} \mathbf{r}_j}{p}.$$

Since for all  $j$ ,

$$\left| \frac{w_j \cdot y_n n w_j}{(1 + w_j \mathbf{r}_j^T (S_{(j,\ell)} - zI)^{-1} \mathbf{r}_j) \cdot (1 + y_n n w_j \widetilde{m}_n(z))} \right| \leq \frac{\kappa^2 y_n}{n},$$

it suffices to show that

$$(3.14) \quad \max_{\ell=1,\dots,n} \max_{j \neq \ell} p^\varepsilon |\zeta_{j,\ell}| \rightarrow 0 \quad \text{almost surely.}$$

To prove this, recall that  $\mathbf{r}_\ell = \Sigma^{1/2} \mathbf{Z}_\ell$ , by Lemma 9 and the independence between  $\mathbf{Z}_\ell$  and  $\Sigma^{1/2}(S_{(j,\ell)} - zI)^{-1} \Sigma^{1/2}$ , for any  $k \in \mathbb{N}$ ,

$$\begin{aligned} (3.15) \quad & E |\mathbf{Z}_\ell^T \Sigma^{1/2} (S_{(j,\ell)} - zI)^{-1} \Sigma^{1/2} \mathbf{Z}_\ell - \text{tr}(\Sigma^{1/2} (S_{(j,\ell)} - zI)^{-1} \Sigma^{1/2})|^{2k} \\ & \leq C_k p^k \cdot E(\|\Sigma^{1/2} (S_{(j,\ell)} - zI)^{-1} \Sigma^{1/2}\|^{2k}) \\ & \leq \frac{C_k p^k \cdot E\|\Sigma\|^{2k}}{v^{2k}} \leq \frac{C p^k p^{2\delta k}}{v^{2k}}, \end{aligned}$$

where in the last line we used Lemma 5 and assumption (A.vii). Hence, for any  $\varepsilon < 1/2 - \delta$ , choosing  $k > 3/(1 - 2\delta - 2\varepsilon)$  and using Borel–Cantelli again, we get

$$(3.16) \quad \begin{aligned} \max_{\ell=1,\dots,n} \max_{j \neq \ell} p^\varepsilon \left| \frac{\mathbf{Z}_\ell^T \Sigma^{1/2} (S_{(j,\ell)} - zI)^{-1} \Sigma^{1/2} \mathbf{Z}_\ell}{p} \right. \\ \left. - \frac{\text{tr}(\Sigma^{1/2} (S_{(j,\ell)} - zI)^{-1} \Sigma^{1/2})}{p} \right| \rightarrow 0. \end{aligned}$$

Furthermore, by Lemma 2 and assumption (A.vii), recall that  $\tilde{m}_n(z) = \text{tr}(\Sigma^{1/2}(S - zI)^{-1}\Sigma^{1/2})/p$ ,

$$(3.17) \quad \begin{aligned} & \max_{\ell=1,\dots,n} \max_{j \neq \ell} \left| \frac{1}{p} \text{tr}(\Sigma^{1/2}(S_{(j,\ell)} - zI)^{-1}\Sigma^{1/2}) - \tilde{m}_n(z) \right| \\ & \leq 2 \frac{\|\Sigma\|}{pv} \leq \frac{Cp^\delta}{pv}. \end{aligned}$$

The convergence (3.14) follows.  $\square$

We now continue the proof of the theorem. Recall that  $\mathbf{r}_\ell = \Sigma^{1/2}\mathbf{Z}_\ell$ , and  $\mathbf{Z}_\ell$  consists of i.i.d. random variables with finite moments of all orders. By Lemma 4(ii) and (3.6),

$$(3.18) \quad \begin{aligned} & \max_{\ell=1,\dots,n} \frac{|\mathbf{r}_\ell^T(S_{(\ell)} - zI)^{-1}(M_n\Sigma + I)^{-1}\mathbf{r}_\ell - \mathbf{r}_\ell^T(S_{(\ell)} - zI)^{-1}(M_{(\ell)}\Sigma + I)^{-1}\mathbf{r}_\ell|}{p} \\ & \leq \max_{\ell=1,\dots,n} \frac{|M_{(\ell)} - M_n(z)| \cdot |\mathbf{r}_\ell|^2 \cdot \|(S_{(\ell)} - zI)^{-1}\| \cdot \|\Sigma\|}{p} \\ & \leq \max_{\ell=1,\dots,n} \frac{|M_{(\ell)} - M_n(z)| \cdot Cp^\delta}{v} \cdot \frac{|\mathbf{r}_\ell|^2}{p} \rightarrow 0, \end{aligned}$$

where in the last line we used Lemma 5, assumption (A.vii), the assumption that  $\delta < 1/6$  (and hence  $\delta < 1/2 - \delta$ ) and (3.11), and (3.3).

Furthermore, similar to (3.15), by Lemma 9 and the independence between  $\mathbf{Z}_\ell$  and  $\Sigma^{1/2}(S_{(\ell)} - zI)^{-1}(M_{(\ell)}\Sigma + I)^{-1}\Sigma^{1/2}$ , for any  $k \in \mathbb{N}$ ,

$$\begin{aligned} & E|\mathbf{Z}_\ell^T \Sigma^{1/2}(S_{(\ell)} - zI)^{-1}(M_{(\ell)}\Sigma + I)^{-1}\Sigma^{1/2}\mathbf{Z}_\ell \\ & \quad - \text{tr}(\Sigma^{1/2}(S_{(\ell)} - zI)^{-1}(M_{(\ell)}\Sigma + I)^{-1}\Sigma^{1/2})|^{2k} \\ & \leq C_k p^k \cdot E(\|\Sigma^{1/2}(S_{(\ell)} - zI)^{-1}(M_{(\ell)}\Sigma + I)^{-1}\Sigma^{1/2}\|^{2k}) \\ & \leq \frac{C_k p^k \cdot E\|\Sigma\|^{2k}}{v^{2k}} \leq \frac{Cp^k p^{2\delta k}}{v^{2k}}, \end{aligned}$$

where in the last line we use Lemmas 5, 3 and (3.6), and assumption (A.vii). Hence, choosing  $k > 2/(1 - 2\delta)$  and using Borel–Cantelli again, we get

$$(3.19) \quad \begin{aligned} & \max_{\ell=1,\dots,n} \left| \frac{\mathbf{Z}_\ell^T \Sigma^{1/2}(S_{(\ell)} - zI)^{-1}(M_{(\ell)}\Sigma + I)^{-1}\Sigma^{1/2}\mathbf{Z}_\ell}{p} \right. \\ & \quad \left. - \frac{\text{tr}(\Sigma^{1/2}(S_{(\ell)} - zI)^{-1}(M_{(\ell)}\Sigma + I)^{-1}\Sigma^{1/2})}{p} \right| \rightarrow 0. \end{aligned}$$

Furthermore, by Lemmas 4(i), 3 and (3.6), the assumption that  $\delta < 1/6$  (and hence  $2\delta < 1/2 - \delta$ ) and (3.11), and assumption (A.vii),

$$\begin{aligned}
 (3.20) \quad & \max_{\ell=1,\dots,n} \left| \frac{1}{p} \operatorname{tr}(\Sigma^{1/2}(S_{(\ell)} - zI)^{-1}(M_{(\ell)}\Sigma + I)^{-1}\Sigma^{1/2}) \right. \\
 & \quad \left. - \frac{1}{p} \operatorname{tr}(\Sigma^{1/2}(S_{(\ell)} - zI)^{-1}(M_n\Sigma + I)^{-1}\Sigma^{1/2}) \right| \\
 & \leq \max_{\ell=1,\dots,n} |M_{(\ell)} - M_n| \cdot \|(S_{(\ell)} - zI)^{-1}\| \cdot \|\Sigma\|^2 \\
 & \leq \max_{\ell=1,\dots,n} |M_{(\ell)} - M_n| \cdot \frac{Cp^{2\delta}}{v} \rightarrow 0.
 \end{aligned}$$

Finally, similar to (3.17), by Lemmas 2 and 3, and assumption (A.vii),

$$\begin{aligned}
 (3.21) \quad & \max_{\ell=1,\dots,n} \left| \frac{1}{p} \operatorname{tr}(\Sigma^{1/2}(S_{(\ell)} - zI)^{-1}(M_n\Sigma + I)^{-1}\Sigma^{1/2}) \right. \\
 & \quad \left. - \frac{1}{p} \operatorname{tr}(\Sigma^{1/2}(S - zI)^{-1}(M_n\Sigma + I)^{-1}\Sigma^{1/2}) \right| \\
 & \leq \frac{\|(M_n\Sigma + I)^{-1}\| \cdot \|\Sigma\|}{pv} \leq \frac{Cp^\delta}{pv} \rightarrow 0.
 \end{aligned}$$

Combining (3.18), (3.19), (3.20) and (3.21), we see that (3.10), and hence (3.7) holds.

Now we are ready to show that  $\{m_n(z)\}$  admits at most one limit.

CLAIM 2. *Suppose that  $m_{n_k}(z)$  converges to  $m(z)$ , then*

$$(3.22) \quad M_{n_k}(z) \rightarrow M(z) := -\frac{1}{z} \int_0^1 \frac{w_s}{1 + y\tilde{m}(z)w_s} ds \neq 0,$$

where  $\tilde{m}(z)$  is the unique solution in  $Q_1 = \{z \in \mathbb{C} : \operatorname{Re}(z) \geq 0, \operatorname{Im}(z) \geq 0\}$  to the following equation:

$$(3.23) \quad \int_0^1 \frac{1}{1 + y\tilde{m}(z)w_s} ds = 1 - y(1 + zm(z)).$$

PROOF. Writing

$$S - zI + zI = \sum_{\ell=1}^n w_\ell \mathbf{r}_\ell \mathbf{r}_\ell^T,$$

right-multiplying both sides by  $(S - zI)^{-1}$  and using (3.8) we get

$$I + z(S - zI)^{-1} = \sum_{\ell=1}^n \frac{w_\ell \mathbf{r}_\ell \mathbf{r}_\ell^T (S_{(\ell)} - zI)^{-1}}{1 + w_\ell \mathbf{r}_\ell^T (S_{(\ell)} - zI)^{-1} \mathbf{r}_\ell}.$$

Taking trace and dividing by  $n$  we get

$$y_n + zy_n m_n(z) = 1 - \frac{1}{n} \sum_{\ell=1}^n \frac{1}{1 + w_\ell \mathbf{r}_\ell^T (S_{(\ell)} - zI)^{-1} \mathbf{r}_\ell},$$

where, recall that,  $m_n(z) = \text{tr}((S - zI)^{-1})/p$  is the Stieltjes transform of  $F^S$ . Hence, if  $m_{n_k}(z) \rightarrow m(z)$ , then

$$\begin{aligned} & \frac{1}{n_k} \sum_{\ell=1}^{n_k} \frac{1}{1 + y_{n_k} n_k w_\ell \cdot \mathbf{r}_\ell^T (S_{(\ell)} - zI)^{-1} \mathbf{r}_\ell / p_k} \\ (3.24) \quad &= \frac{1}{n_k} \sum_{\ell=1}^{n_k} \frac{1}{1 + w_\ell \mathbf{r}_\ell^T (S_{(\ell)} - zI)^{-1} \mathbf{r}_\ell} \\ &= 1 - y_{n_k} (1 + z m_{n_k}(z)) \rightarrow 1 - y(1 + z m(z)). \end{aligned}$$

However, by the same arguments for (3.16) and (3.17) we have

$$(3.25) \quad \max_{\ell=1, \dots, n} \left| \frac{\mathbf{r}_\ell^T (S_{(\ell)} - zI)^{-1} \mathbf{r}_\ell}{p} - \frac{\text{tr}(\Sigma^{1/2} (S_{(\ell)} - zI)^{-1} \Sigma^{1/2})}{p} \right| \rightarrow 0$$

and

$$(3.26) \quad \max_{\ell=1, \dots, n} \left| \frac{\text{tr}(\Sigma^{1/2} (S_{(\ell)} - zI)^{-1} \Sigma^{1/2})}{p} - \tilde{m}_n(z) \right| \rightarrow 0,$$

where, recall that  $\tilde{m}_n(z) = \text{tr}(\Sigma^{1/2} (S - zI)^{-1} \Sigma^{1/2})/p$ , which belongs to  $Q_1$  by Lemma 7. Then by (3.24), assumption (A.v) and Lemma 8,  $\tilde{m}_{n_k}(z)$  must also converge, and the limit, denoted by  $\tilde{m}(z) \in Q_1$ , must be the unique solution in  $Q_1$  to the equation (3.23). Now by (3.12), (3.13) and assumption (A.v), we get the convergence for  $M_{n_k}(z)$  in the claim. That  $M(z) \neq 0$  follows from the expression and that  $\tilde{m}(z) \in Q_1$ .  $\square$

We now continue the proof of the theorem. By the convergence of  $F^{\Sigma_p}$  to  $H$  and the previous claim,

$$\frac{\text{tr}((-zM_{n_k}(z)\Sigma - zI)^{-1})}{p} \rightarrow -\frac{1}{z} \int_{\tau \in \mathbb{R}} \frac{1}{\tau M(z) + 1} dH(\tau).$$

But (3.7) implies that

$$(3.27) \quad m(z) = -\frac{1}{z} \int_{\tau \in \mathbb{R}} \frac{1}{\tau M(z) + 1} dH(\tau).$$

Observing that  $M(z) \neq 0$ ,  $\text{Re}(M(z)) \geq 0$ , and  $H$  is not a delta measure at 0, we obtain that  $|m(z)| < 1/|z|$ . Hence  $1 + zm(z) \neq 0$ , and by (3.23),  $\tilde{m}(z) \neq 0$ . Based on this, we can get another expression for  $M(z)$ , as follows.

By (3.22), we have

$$\begin{aligned}
 M(z) &= -\frac{1}{z} \int_0^1 \frac{w_s}{1 + y\tilde{m}(z)w_s} ds \\
 &= -\frac{1}{z} \cdot \frac{1}{y\tilde{m}(z)} \cdot \left(1 - \int_0^1 \frac{1}{1 + y\tilde{m}(z)w_s} ds\right) \\
 (3.28) \quad &= -\frac{1}{z} \cdot \frac{1}{y\tilde{m}(z)} \cdot (1 - (1 - y(1 + zm(z)))) \\
 &= -\frac{1}{z} \cdot \frac{1 + zm(z)}{\tilde{m}(z)},
 \end{aligned}$$

where in the third line we used the definition (3.23) of  $\tilde{m}(z)$ .

We can then derive another formula for  $\tilde{m}(z)$ . By (3.27),

$$1 + zm(z) = 1 - \int_{\tau \in \mathbb{R}} \frac{1}{\tau M(z) + 1} dH(\tau) = M(z) \int_{\tau \in \mathbb{R}} \frac{\tau}{\tau M(z) + 1} dH(\tau)$$

by using that  $H$  is a probability distribution. Dividing both sides by  $-z\tilde{m}(z) (\neq 0)$  and using (3.28) yield

$$M(z) = -\frac{(1 + zm(z))}{z\tilde{m}(z)} = -\frac{M(z) \int_{\tau \in \mathbb{R}} \tau / (\tau M(z) + 1) dH(\tau)}{z\tilde{m}(z)},$$

and hence since  $M(z) \neq 0$ ,

$$(3.29) \quad \tilde{m}(z) = -\frac{1}{z} \int_{\tau \in \mathbb{R}} \frac{\tau}{\tau M(z) + 1} dH(\tau).$$

Observe that by Lemma 7 and (3.6), for any  $n$ , both  $m_n(z)$  and  $M_n(z)$  belong to  $Q_1$ , hence so do  $m(z)$  and  $M(z)$ . We proceed to show that for those  $z = iv$  with  $v$  sufficiently large, there is at most one triple  $(m(z), M(z), \tilde{m}(z)) \in Q_1 \times Q_1 \times Q_1$  that solves the equations (3.27), (3.22) and (3.29). In fact, if there are two different triples  $(m_i(z), M_i(z), \tilde{m}_i(z)), i = 1, 2$ , both satisfying (3.27), (3.22) and (3.29). Then necessarily,  $M_1(z) \neq M_2(z)$  and  $\tilde{m}_1(z) \neq \tilde{m}_2(z)$ . Now by (3.22),

$$M_1(z) - M_2(z) = -\frac{1}{z} \cdot y(\tilde{m}_2(z) - \tilde{m}_1(z)) \int_0^1 \frac{w_s^2}{(1 + y\tilde{m}_1(z)w_s)(1 + y\tilde{m}_2(z)w_s)} ds$$

by (3.29),

$$\tilde{m}_1(z) - \tilde{m}_2(z) = -\frac{1}{z} (M_2(z) - M_1(z)) \int_{\tau \in \mathbb{R}} \frac{\tau^2}{(\tau M_1(z) + 1)(\tau M_2(z) + 1)} dH(\tau).$$

Therefore,

$$\begin{aligned}
 1 &= \frac{y}{z^2} \int_0^1 \frac{w_s^2}{(1 + y\tilde{m}_1(z)w_s)(1 + y\tilde{m}_2(z)w_s)} ds \\
 (3.30) \quad &\times \int_{\tau \in \mathbb{R}} \frac{\tau^2}{(\tau M_1(z) + 1)(\tau M_2(z) + 1)} dH(\tau).
 \end{aligned}$$

However, since  $(M_i(z), \tilde{m}_i(z)) \in Q_1 \times Q_1$ ,  $i = 1, 2$ ,

$$\left| \int_0^1 \frac{w_s^2}{(1 + y\tilde{m}_1(z)w_s)(1 + y\tilde{m}_2(z)w_s)} ds \right| \leq \int_0^1 w_s^2 ds < \infty$$

and

$$\left| \int_{\tau \in \mathbb{R}} \frac{\tau^2}{(\tau M_1(z) + 1)(\tau M_2(z) + 1)} dH(\tau) \right| \leq \int_{\tau \in \mathbb{R}} \tau^2 dH(\tau) < \infty.$$

Hence, for  $z = iv$  with  $v$  sufficiently large, (3.30) cannot be true.

It remains to verify (3.1), that is,  $\lim_{v \rightarrow \infty} iv \cdot m(iv) = -1$ . In fact, using (3.27) we get that

$$(3.31) \quad iv \cdot m(iv) = - \int_{\tau \in \mathbb{R}} \frac{1}{1 + \tau M(iv)} dH(\tau).$$

Since  $\operatorname{Re}(M(iv)) \geq 0$ ,  $|1/(1 + \tau M(iv))| \leq 1$  for all  $\tau \geq 0$ . Moreover, by (3.22) and that  $\operatorname{Re}(\tilde{m}(z)) \geq 0$ ,  $|M(iv)| \leq 1/v \cdot \int_0^1 w_s ds$ , hence by the dominated convergence theorem, the right-hand side of (3.31) converges to  $-1$  as  $v \rightarrow \infty$ .  $\square$

**3.4. Proof of Theorem 2.** The TVARCV matrix has the form of weighted sample covariance matrices as studied in Theorem 1; however, assumption (A.vi) therein is not satisfied, and we need another proof.

Theorem 2 is a direct consequence of the following two convergence results.

**PROPOSITION 7.** *Under assumption (C.iv), namely, suppose that*

$$\lim_{p \rightarrow \infty} \operatorname{tr}(\Sigma_p)/p = \theta,$$

*then, almost surely,  $\lim_{p \rightarrow \infty} \operatorname{tr}(\Sigma_p^{\text{RCV}})/p = \theta$ .*

The proof is given in the supplemental article [Zheng and Li (2011)].

Next, recall that  $\check{\Sigma}_p$  and  $\tilde{\Sigma}_p$  are defined by (2.12) and (2.11), respectively.

**PROPOSITION 8.** *Under the assumptions of Theorem 2, both  $F^{\check{\Sigma}_p}$  and  $F^{\tilde{\Sigma}_p}$  converge almost surely.  $F^{\check{\Sigma}_p}$  converges to  $\check{H}$  defined by*

$$(3.32) \quad \check{H}(x) = H(\theta x) \quad \text{for all } x \geq 0.$$

*The LSD  $\tilde{F}$  of  $\tilde{\Sigma}_p$  is determined by  $\check{H}$  in that its Stieltjes transform  $m_{\tilde{F}}(z)$  satisfies the equation*

$$m_{\tilde{F}}(z) = \int_{\tau \in \mathbb{R}} \frac{1}{\tau(1 - y(1 + zm_{\tilde{F}}(z))) - z} d\check{H}(\tau).$$

This can be proved in very much the same way as Theorem 1, by working with Stieltjes transforms. However, a much simpler and transparent proof is as follows.



PROOF OF PROPOSITION 8. The convergence of  $F^{\check{\Sigma}_p}$  is obvious since

$$F^{\check{\Sigma}_p}(x) = F^{\Sigma_p}(\text{tr}(\Sigma_p)/p \cdot x) \quad \text{for all } x \geq 0.$$

We now show the convergence of  $F^{\check{\Sigma}_p}$ . As in the proof of Theorem 1, for notational ease, we shall sometimes omit the superscript  $p$  in the arguments below: thus, we write  $\boldsymbol{\mu}_t$  instead of  $\boldsymbol{\mu}_t^{(p)}$ ,  $\gamma_t$  instead of  $\gamma_t^{(p)}$ ,  $\Lambda$  instead of  $\Lambda^{(p)}$ , etc.

First, note that

$$\Delta \mathbf{X}_\ell = \int_{\tau_{n,\ell-1}}^{\tau_{n,\ell}} \boldsymbol{\mu}_t dt + \Lambda \cdot \int_{\tau_{n,\ell-1}}^{\tau_{n,\ell}} \gamma_t d\mathbf{W}_t := \sqrt{\int_{\tau_{n,\ell-1}}^{\tau_{n,\ell}} \gamma_t^2 dt} \cdot (\mathbf{v}_\ell + \Lambda \cdot \mathbf{Z}_\ell),$$

where

$$\mathbf{v}_\ell = \begin{pmatrix} v_\ell^{(1)} \\ \vdots \\ v_\ell^{(p)} \end{pmatrix} = \frac{\int_{\tau_{n,\ell-1}}^{\tau_{n,\ell}} \boldsymbol{\mu}_t dt}{\sqrt{\int_{\tau_{n,\ell-1}}^{\tau_{n,\ell}} \gamma_t^2 dt}} \quad \text{and} \quad \mathbf{Z}_\ell = \begin{pmatrix} Z_\ell^{(1)} \\ \vdots \\ Z_\ell^{(p)} \end{pmatrix} = \frac{\int_{\tau_{n,\ell-1}}^{\tau_{n,\ell}} \gamma_t d\mathbf{W}_t}{\sqrt{\int_{\tau_{n,\ell-1}}^{\tau_{n,\ell}} \gamma_t^2 dt}}.$$

By performing an orthogonal transformation if necessary, without loss of generality, we may assume that the index set  $\mathcal{I}_p \subset \{1, \dots, \eta_p\}$ . Then by assumptions (C.ii) and (C.ix), for  $j > \eta_p$ ,  $Z_\ell^{(j)}$  are i.i.d.  $N(0, 1)$ . Write  $\mathbf{U}_\ell = (Z_\ell^{(1)}, \dots, Z_\ell^{(\eta_p)})^T$  and  $\mathbf{D}_\ell = (Z_\ell^{(\eta_p+1)}, \dots, Z_\ell^{(p)})^T$ . With the above notation,  $\check{\Sigma}_p$  can be rewritten as

$$(3.33) \quad \check{\Sigma}_p = y_n \sum_{\ell=1}^n \frac{\Delta \mathbf{X}_\ell (\Delta \mathbf{X}_\ell)^T}{|\Delta \mathbf{X}_\ell|^2} = y_n \sum_{\ell=1}^n \frac{(\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell)(\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell)^T}{|\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell|^2}.$$

By assumptions (C.i), (C.ii) and (C.ix), there exists  $C > 0$  such that  $|v_\ell^{(j)}| \leq C/\sqrt{n}$  for all  $j$  and  $\ell$ , hence  $|\mathbf{v}_\ell|$ 's are uniformly bounded. We will show that

$$(3.34) \quad \begin{aligned} & \max_{\ell=1, \dots, n} ||\Lambda \mathbf{Z}_\ell|^2/p - 1| \\ &= \max_{\ell=1, \dots, n} |\mathbf{Z}_\ell^T \check{\Sigma}_p \mathbf{Z}_\ell/p - 1| \rightarrow 0 \quad \text{almost surely,} \end{aligned}$$

which clearly implies that

$$(3.35) \quad \max_{\ell=1, \dots, n} ||\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell|^2/p - 1| \rightarrow 0 \quad \text{almost surely.}$$

To prove (3.34), write

$$\check{\Sigma}_p = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix},$$

where  $A, B$  and  $C$  are  $\eta_p \times \eta_p, \eta_p \times (n - \eta_p)$  and  $(n - \eta_p) \times (n - \eta_p)$  matrices, respectively. Then

$$\mathbf{Z}_\ell^T \check{\Sigma}_p \mathbf{Z}_\ell = \mathbf{U}_\ell^T A \mathbf{U}_\ell + 2\mathbf{D}_\ell^T B^T \mathbf{U}_\ell + \mathbf{D}_\ell^T C \mathbf{D}_\ell.$$

By a well-known fact about the spectral norm,

$$\|A\| \leq \|\check{\Sigma}_p\|, \quad \|B\| \leq \|\check{\Sigma}_p\| \quad \text{and} \quad \|C\| \leq \|\check{\Sigma}_p\|.$$

In particular, by assumptions (C.ii), (C.vi) and (C.vii),

$$0 \leq \text{tr}(A) \leq \eta_p \cdot \|\check{\Sigma}_p\| \leq Cp^{\delta_1 + \delta_2} = o(p),$$

hence  $\text{tr}(C)/p = (\text{tr}(\check{\Sigma}_p) - \text{tr}(A))/p \rightarrow 1$ . Now using the fact that  $\mathbf{D}_\ell$  consists of i.i.d. standard normals and by the same proof as that for (3.4) we get

$$(3.36) \quad \max_{\ell=1, \dots, n} |\mathbf{D}_\ell^T C \mathbf{D}_\ell / p - 1| \rightarrow 0 \quad \text{almost surely.}$$

To complete the proof of (3.34), it then suffices to show that

$$\max_{\ell=1, \dots, n} \frac{|\mathbf{U}_\ell^T A \mathbf{U}_\ell|}{p} \rightarrow 0 \quad \text{and} \quad \max_{\ell=1, \dots, n} \frac{|\mathbf{D}_\ell^T B^T \mathbf{U}_\ell|}{p} \rightarrow 0 \quad \text{almost surely.}$$

We shall only prove the first convergence; the second one can be proved similarly. We have

$$(3.37) \quad |\mathbf{U}_\ell^T A \mathbf{U}_\ell| \leq \|A\| \cdot |\mathbf{U}_\ell|^2 \leq C_5 p^{\delta_2} \cdot |\mathbf{U}_\ell|^2.$$

Observe that for all  $1 \leq i \leq \eta_p$ , by assumption (C.ii),

$$|Z_\ell^{(i)}|^2 = \frac{|\int_{\tau_{n,\ell-1}}^{\tau_{n,\ell}} \gamma_t d\mathbf{W}_t|^2}{\int_{\tau_{n,\ell-1}}^{\tau_{n,\ell}} \gamma_t^2 dt} \leq \frac{C_2^2}{\Delta \tau_{n,\ell}} \cdot \left| \int_{\tau_{n,\ell-1}}^{\tau_{n,\ell}} \gamma_t d\mathbf{W}_t \right|^2.$$

By the Burkholder–Davis–Gundy inequality, we then get that for any  $k \in \mathbb{N}$ , there exists  $\lambda_k > 0$  such that

$$(3.38) \quad E|Z_\ell^{(i)}|^{2k} \leq \lambda_k C_2^{4k}.$$

Now we are ready to show that  $\max_{\ell=1, \dots, n} |\mathbf{U}_\ell^T A \mathbf{U}_\ell|/p \rightarrow 0$ . In fact, for any  $\varepsilon > 0$ , for any  $k \in \mathbb{N}$ , by Markov's inequality, (3.37), Hölder's inequality and (3.38),

$$\begin{aligned} P\left(\max_{\ell=1, \dots, n} |\mathbf{U}_\ell^T A \mathbf{U}_\ell| \geq p\varepsilon\right) &\leq \sum_{\ell=1}^n P(|\mathbf{U}_\ell^T A \mathbf{U}_\ell| \geq p\varepsilon) \\ &\leq \sum_{\ell=1}^n \frac{E|\mathbf{U}_\ell^T A \mathbf{U}_\ell|^k}{p^k \varepsilon^k} \\ &\leq \sum_{\ell=1}^n \frac{C_5^k p^{k\delta_2} \cdot [(\eta_p \cdot \lambda_k C_2^{4k}) \cdot \eta_p^{k-1}]}{p^k \varepsilon^k} \\ &\leq Cp^{1+k\delta_2+k\delta_1-k}. \end{aligned}$$

By assumption (C.vii),  $\delta_1 + \delta_2 < 1/2 < 1$ , hence by choosing  $k$  to be large enough, the right hand side will be summable in  $p$ , hence by Borel–Cantelli, almost surely,  $\max_{\ell=1,\dots,n} |\mathbf{U}_\ell^T \mathbf{A} \mathbf{U}_\ell|/p \rightarrow 0$ .

We now get back to  $\tilde{\Sigma}_p$  as in (3.33). By (3.35), for any  $\varepsilon > 0$ , almost surely, for all  $n$  sufficiently large, for all  $\ell = 1, \dots, n$ ,

$$p(1 - \varepsilon) \leq |\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell|^2 \leq p(1 + \varepsilon).$$

Hence, almost surely, for all  $n$  sufficiently large,

$$\frac{1}{1 + \varepsilon} \tilde{S}_p \leq \tilde{\Sigma}_p = y_n \sum_{\ell=1}^n \frac{(\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell)(\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell)^T}{|\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell|^2} \leq \frac{1}{1 - \varepsilon} \tilde{S}_p,$$

where  $\tilde{S}_p = 1/n \cdot \sum_{1 \leq \ell \leq n} (\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell)(\mathbf{v}_\ell + \Lambda \mathbf{Z}_\ell)^T$ . Hence, by Weyl’s Monotonicity theorem, for any  $x \geq 0$ ,

$$(3.39) \quad F^{\tilde{S}_p}((1 + \varepsilon)x) \geq F^{\tilde{\Sigma}_p}(x) \geq F^{\tilde{S}_p}((1 - \varepsilon)x).$$

Next, by Lemma 1,  $\tilde{S}_p$  has the same LSD as  $S_p := 1/n \sum_{1 \leq \ell \leq n} \Lambda \mathbf{Z}_\ell (\mathbf{Z}_\ell)^T \Lambda^T$ . Moreover, by using the same trick as in the beginning of the proof of Theorem 1,  $F^{S_p}$  has the same limit as  $F^{S'_p}$ , where  $S'_p = 1/n \sum_{1 \leq \ell \leq n} \Lambda \tilde{\mathbf{Z}}_\ell (\tilde{\mathbf{Z}}_\ell)^T \Lambda^T$ , and  $\tilde{\mathbf{Z}}_\ell$  consists of i.i.d. standard normals. For  $F^{S'_p}$ , it follows easily from Proposition 1 that it converges to  $\tilde{F}$ . Moreover, by Theorems 1.1 and 2.1 in Silverstein and Choi (1995),  $\tilde{F}$  is differentiable and in particular continuous at all  $x > 0$ . It follows from (3.39) that  $F^{\tilde{\Sigma}_p}$  must also converge to  $\tilde{F}$ .  $\square$

**4. Simulation studies.** In this section, we present some simulation studies to illustrate the behavior of ESDs of RCV and TVARCV matrices. In particular, we show that the ESDs of RCV matrices that have the same targeting ICV matrix  $\Sigma_p$  can be quite different from each other, depending on the time variability of the covolatility process. Our proposed estimator, the TVARCV matrix  $\hat{\Sigma}_p$ , in contrast, has a very stable ESD.

We use in particular a reference curve which is the Marčenko–Pastur law. The reason we compare the ESDs of RCV and TVARCV matrices with the Marčenko–Pastur law is that the Marčenko–Pastur law is the LSD of  $\Sigma_p^{\text{RCV}^0}$  defined in (1.6), which is the RCV matrix estimated from sample paths of constant volatility that has the same targeting ICV matrix as  $\Sigma_p^{\text{RCV}}$ . As we will see soon in the following two subsections, when the covolatility process is time varying, the ESD of RCV matrix can be very different from the Marčenko–Pastur law, while the ESD of TVARCV matrix always matches the Marčenko–Pastur law very well.

In the simulation below, we assume that  $\Lambda = I$ , or in other words,  $\mathbf{X}_t$  satisfies (2.1) with  $\gamma_t$  a deterministic (scalar) process, and  $\mathbf{W}_t$  a  $p$ -dimensional

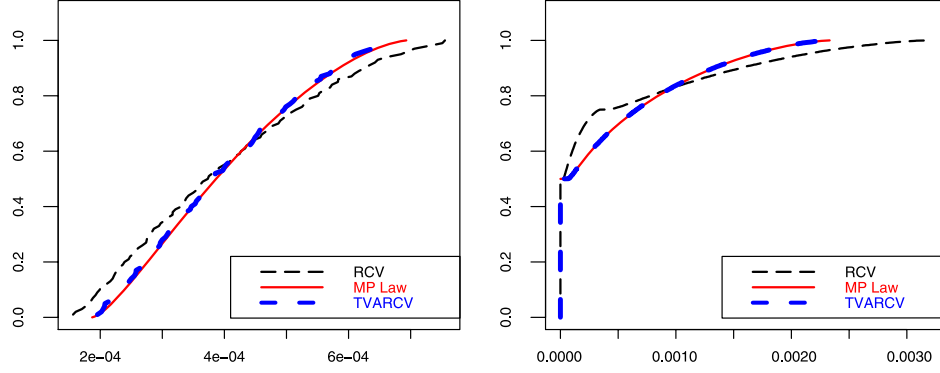


FIG. 1. Left panel:  $p = 100$ ,  $n = 1,000$ ; right panel:  $p = 2,000$ ,  $n = 1,000$ .

standard Brownian motion. The observation times are taken to be equidistant:  $\tau_{n,\ell} = \ell/n$ ,  $\ell = 0, 1, \dots, n$ .

We present simulation results of two different designs: one when  $\gamma_t$  is piecewise constant, the other when  $\gamma_t$  is continuous (and non-constant). In both cases, we compare the ESDs of the RCV and TVARCV matrices. Results for different dimension  $p$  and observation frequency  $n$  are reported.

In all the figures below, we use red solid lines to represent the LSDs of  $\Sigma^{\text{RCV}_0}$  given by the Marčenko–Pastur law, black dashed line to represent the ESDs of RCV matrices, blue bold longdashed line to represent the ESDs of TVARCV matrices.

*4.1. Design I, piecewise constants.* We first consider the case when the volatility path follows piecewise constants. More specifically, we take  $\gamma_t$  to be

$$\gamma_t = \begin{cases} \sqrt{0.0007}, & t \in [0, 1/4) \cup [3/4, 1], \\ \sqrt{0.0001}, & t \in [1/4, 3/4). \end{cases}$$

In Figure 1, we compare the ESDs of RCV and TVARCV matrices for different pairs of  $p$  and  $n$ , with the LSD of  $\Sigma^{\text{RCV}_0}$  given by the Marčenko–Pastur law as reference.

We see from Figure 1 that:

- the ESDs of RCV matrices are very different from the LSD given by the Marčenko–Pastur law (the LSD of  $\Sigma^{\text{RCV}_0}$ );
- the ESDs of TVARCV matrices follow the LSD given by the Marčenko–Pastur law very well, for both pairs of  $p$  and  $n$ , even when  $p$  is small compared with  $n$ .

In fact, the dependence of the ESD of RCV matrix on the time variability of covolatility process can be seen more clearly from Figure 2, where we

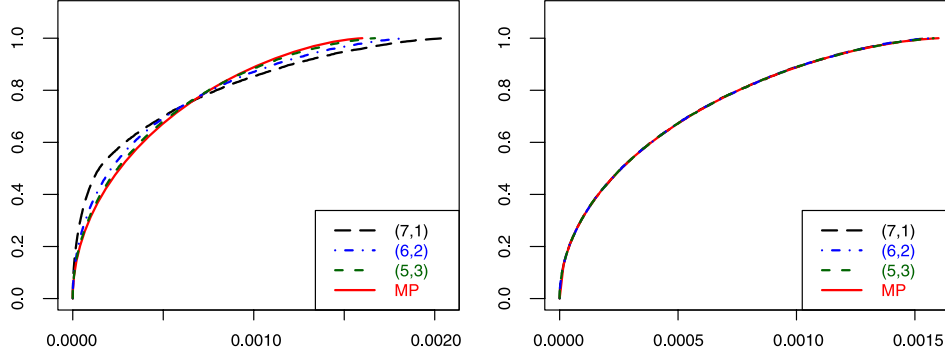


FIG. 2. Comparisons, different values of piecewise constants  $(a, b)$  as shown in the legend, which are such that the targeting ICV matrix is the same; the red solid curve is the Marčenko–Pastur law (the LSD of  $\Sigma^{\text{RCV}^0}$ ).  $p$  and  $n$  are both taken to be 1,000. Left panel: RCV; right panel: TVARCV.

consider the same design but different values for  $\gamma_t$ :

$$\gamma_t = \begin{cases} a^{1/2} \times 10^{-2}, & t \in [0, 1/4) \cup [3/4, 1], \\ b^{1/2} \times 10^{-2}, & t \in [1/4, 3/4), \end{cases} \quad \text{where } a + b = 8.$$

We plot the ESDs of RCV and TVARCV matrices for the case when  $p = n = 1,000$ , in the left and right panel, respectively. The curves' corresponding parameters  $(a, b)$  are reported in the legend. Note that since all pairs of  $(a, b)$  have the same summation, in all cases the targeting ICV matrices are the same.

We see clearly from Figure 2 that, the ESDs of RCV matrices can be very different from each other even though the RCV matrices are estimating the same ICV matrix; while for TVARCV matrices, the ESDs are almost identical.

**4.2. Design II, continuous paths.** We illustrate in this subsection the case when the volatility processes have continuous sample paths. In particular, we assume that  $\mathbf{X}_t$  satisfies (2.1) with

$$\gamma_t = \sqrt{0.0009 + 0.0008 \cos(2\pi t)}, \quad t \in [0, 1].$$

We see from Figure 3 similar phenomena as in Design I about the ESDs of RCV and TVARCV matrices for different pairs of  $p$  and  $n$ .

**5. Conclusion and discussions.** We have shown theoretically and via simulation studies that:

- the *limiting spectral distribution* (LSD) of RCV matrix depends not only on that of the ICV matrix, but also on the time-variability of covolatility process;

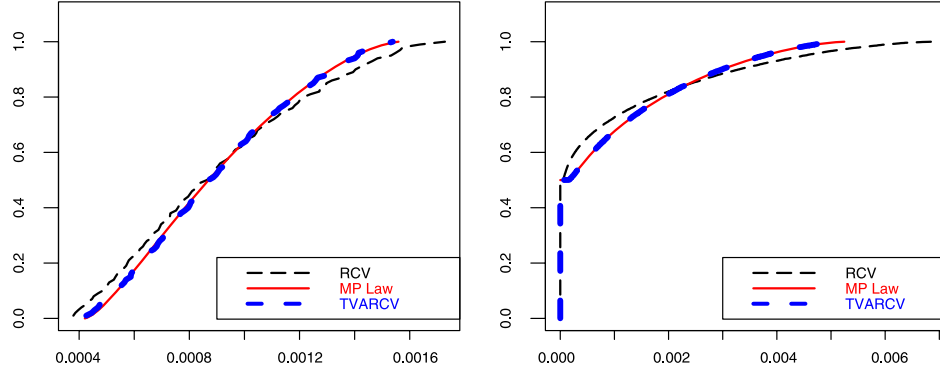


FIG. 3. Left panel:  $p = 100$ ,  $n = 1,000$ ; right panel:  $p = 2,000$ ,  $n = 1,000$ .

- in particular, even with the same targeting ICV matrix, the *empirical spectral distribution* (ESD) of RCV matrix can vary a lot, depending on how the underlying covolatility process evolves over time;
- for a class  $\mathcal{C}$  of processes, our proposed estimator, the *time-variation adjusted realized covariance* (TVARCV) matrix, possesses the following desirable properties as an estimator of the ICV matrix: as long as the targeting ICV matrix is the same, the ESDs of TVARCV matrices estimated from processes with different covolatility paths will be close to each other, sharing a unique limit; moreover, the LSD of TVARCV matrix is related to that of the targeting ICV matrix through the same Marčenko–Pastur equation as in the sample covariance matrix case.

Furthermore, we establish a Marčenko–Pastur type theorem for weighted sample covariance matrices. For a class  $\mathcal{C}$  of processes, we also establish a Marčenko–Pastur type theorem for RCV matrices, which explicitly demonstrates how the time-variability of the covolatility process affects the LSD of RCV matrix.

In practice, for given  $p$  and  $n$ , based on the (observable) ESD of TVARCV matrix, one can use existing algorithms to obtain an estimate of the ESD of ICV matrix, which can then be applied to further applications such as portfolio allocation, risk management, etc.

**Acknowledgments.** We are very grateful to the Editor, the Associate Editor and anonymous referees for their very valuable comments and suggestions.

#### SUPPLEMENTARY MATERIAL

**Supplement to “On the estimation of integrated covariance matrices of high dimensional diffusion processes”** (DOI: [10.1214/11-AOS939SUPP](https://doi.org/10.1214/11-AOS939SUPP); .pdf). This material contains the proof of Proposition 4, a detailed explanation of the second statement in Remark 3, and the proofs of the various lemmas in Section 3.1 and Proposition 7.

## REFERENCES

- ADMATI, A. R. and PFLEIDERER, P. (1988). A theory of intraday patterns: Volume and price variability. *Rev. Financ. Stud.* **1** 3–40.
- ANDERSEN, T. G. and BOLLERSLEV, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance* **4** 115–158.
- ANDERSEN, T. G. and BOLLERSLEV, T. (1998). Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. *J. Finance* **53** 219–265.
- ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and LABYS, P. (2001). The distribution of realized exchange rate volatility. *J. Amer. Statist. Assoc.* **96** 42–55. [MR1952727](#)
- BAI, Z. D. (1999). Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica* **9** 611–677. [MR1711663](#)
- BAI, Z., CHEN, J. and YAO, J. (2010). On estimation of the population spectral distribution from a high-dimensional sample covariance matrix. *Aust. N. Z. J. Stat.* **52** 423–437. [MR2791528](#)
- BAI, Z., LIU, H. and WONG, W.-K. (2009). Enhancement of the applicability of Markowitz’s portfolio optimization by utilizing random matrix theory. *Math. Finance* **19** 639–667. [MR2583523](#)
- BAI, Z. D. and SILVERSTEIN, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.* **26** 316–345. [MR1617051](#)
- BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica* **72** 885–925. [MR2051439](#)
- BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. and SHEPHARD, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J. Econometrics* **162** 149–169.
- EL KAROUI, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Statist.* **36** 2757–2790. [MR2485012](#)
- FAN, J., LI, Y. and YU, K. (2011). Vast volatility matrix estimation using high frequency data for portfolio selection. *J. Amer. Statist. Assoc.* To appear.
- GERONIMO, J. S. and HILL, T. P. (2003). Necessary and sufficient condition that the limit of Stieltjes transforms is a Stieltjes transform. *J. Approx. Theory* **121** 54–60. [MR1962995](#)
- HORN, R. A. and JOHNSON, C. R. (1990). *Matrix Analysis*. Cambridge Univ. Press, Cambridge. Corrected reprint of the 1985 original. [MR1084815](#)
- JACOD, J. and PROTTER, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.* **26** 267–307. [MR1617049](#)
- MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)* **72** 507–536. [MR0208649](#)
- MARKOWITZ, H. (1952). Portfolio selection. *J. Finance* **7** 77–91.
- MARKOWITZ, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investments*. Cowles Foundation for Research in Economics at Yale University, Monograph **16**. Wiley, New York. [MR0103768](#)
- MESTRE, X. (2008). Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Trans. Inform. Theory* **54** 5113–5129. [MR2589886](#)
- SILVERSTEIN, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *J. Multivariate Anal.* **55** 331–339. [MR1370408](#)

- SILVERSTEIN, J. W. and BAI, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.* **54** 175–192. [MR1345534](#)
- SILVERSTEIN, J. W. and CHOI, S.-I. (1995). Analysis of the limiting spectral distribution of large-dimensional random matrices. *J. Multivariate Anal.* **54** 295–309. [MR1345541](#)
- TAO, M., WANG, Y., YAO, Y. and ZOU, J. (2011). Large volatility matrix inference via combining low-frequency and high-frequency approaches. *J. Amer. Statist. Assoc.* **106** 1025–1040.
- WANG, Y. and ZOU, J. (2010). Vast volatility matrix estimation for high-frequency financial data. *Ann. Statist.* **38** 943–978. [MR2604708](#)
- YIN, Y. Q. (1986). Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.* **20** 50–68. [MR0862241](#)
- ZHENG, X. and LI, Y. (2011). Supplement to “On the estimation of integrated covariance matrices of high dimensional diffusion processes.” [DOI:10.1214/11-AOS939SUPP](#).

DEPARTMENT OF INFORMATION SYSTEMS  
BUSINESS STATISTICS AND OPERATIONS MANAGEMENT  
HONG KONG UNIVERSITY OF SCIENCE  
AND TECHNOLOGY  
CLEAR WATER BAY, KOWLOON  
HONG KONG  
E-MAIL: [xhzheng@ust.hk](mailto:xhzheng@ust.hk)  
[yyli@ust.hk](mailto:yyli@ust.hk)